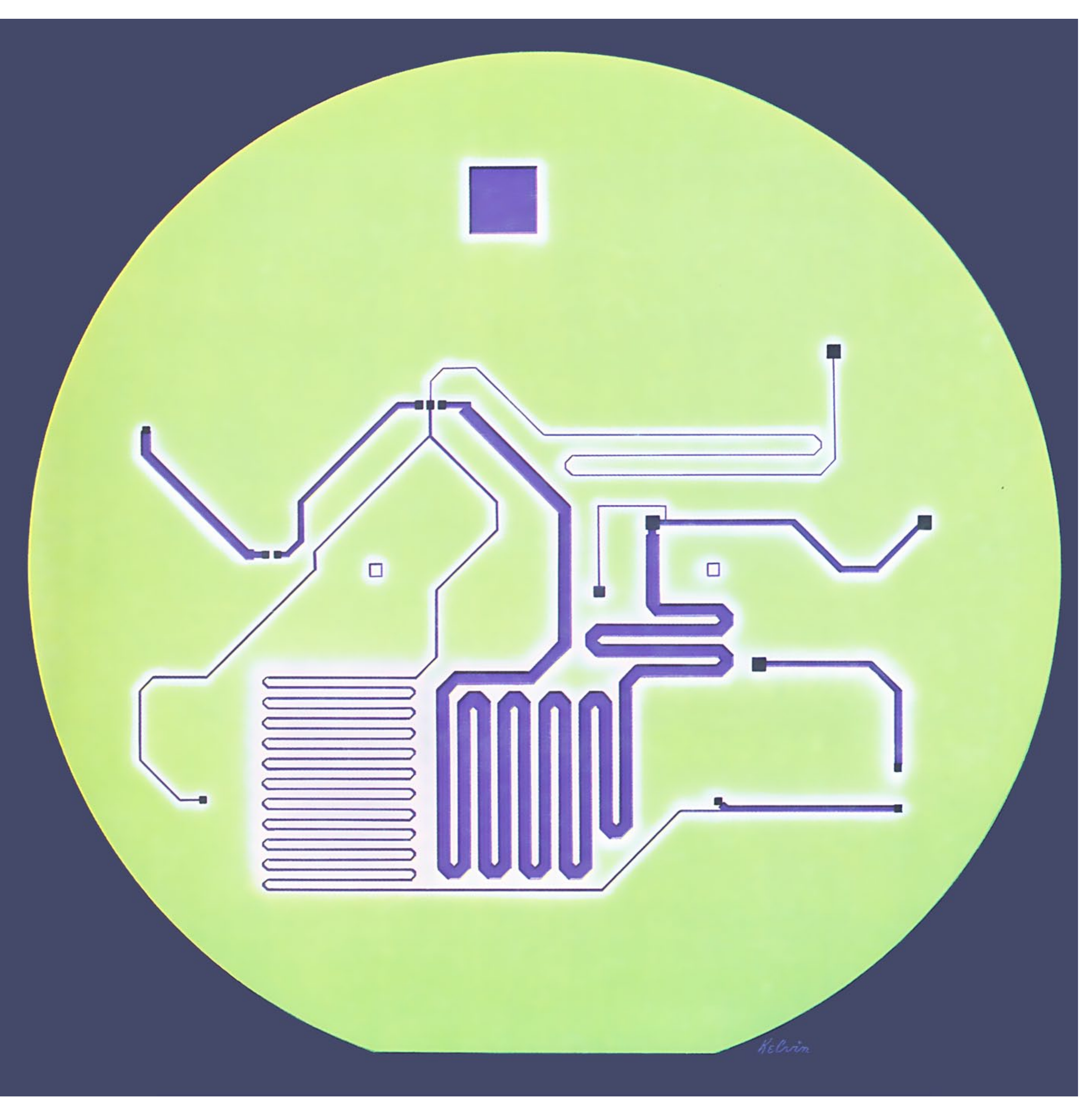
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



DISPOSITIVOS MICROMECANICOS

Junio 1983

350 Ptas.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 81
Junio 1983

INVESTIGACION Y
CIENCIA
Edición en español de SCIENTIFIC AMERICAN

Colaboradores de este número:

Asesoramiento y traducción:

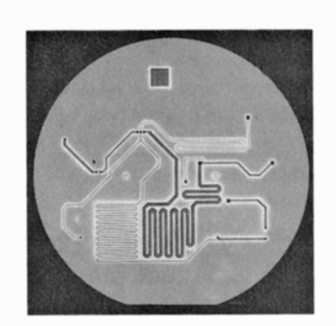
Fernando José Rodilla: *Una zona desnuclearizada en Europa*; Amando García Rodríguez: *Dispositivos micromecánicos de silicio*; August Palanques Mestre: *Estructura de quarks y leptones*; César Casquet Martín: *Fuentes termales en el fondo del océano*; Jesús Cacho Gutiérrez: *Una ventana abierta al cerebro dormido*; Julio Mangas Manjarrés: *Primeros agricultores de la llanura del norte de Europa*; Isabel García Acha y Julio Rodríguez Villanueva: *Señales químicas de agregación en las amebas*; Luis Bou: *Temas metamágicos*; J. Vilardell: *Taller y laboratorio*.

Ciencia y sociedad:

Francisco Gracia

Libros:

J. J. Vázquez, Tomás Santos, Juan Pedro Acordagoicoechea y F. de Diego Calonge



LA PORTADA

En la portada se ilustra una nueva tecnología basada en la "mecanización" de obleas de silicio, similares a las utilizadas para construir los elementos microelectrónicos que se conocen con el nombre de pastillas ("chips") (véase "Dispositivos micromecánicos de silicio", por James B. Angell, Stephen C. Terry y Phillip W. Barth, en este número). Se muestra una oblea de unos ocho centímetros de diámetro en la que se ha construido un cromatógrafo de gases que podría caber en la palma de la mano. Los elementos en forma de serpentín son los canales de los gases. Estos canales se construyen grabando los oportunos surcos en la oblea de silicio y uniéndola posteriormente a una placa de vidrio. El gas procedente de los elementos grabados en la parte delantera de la oblea llega hasta los montados en la parte trasera de dicha oblea a través de una serie de orificios cuadrados que la atraviesan de parte a parte. Los diferentes orificios están cerrados por válvulas; los asientos de las válvulas se graban en la cara posterior de la oblea. La muestra gaseosa a analizar se inyecta a través de una válvula en el canal de un serpentín más ancho y allí se comprime con la ayuda de un pistón.

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 37

Condiciones de suscripción:
España:
Un año (12 números): 3.850 pesetas
Extranjero:
Un año (12 números): 52 U.S. $
Ejemplar atrasado ordinario:
350 pesetas
Ejemplar atrasado extraordinario:
450 pesetas

Distribución para España
Distribuciones de Enlace, S. A.
Ausias March, 49 - Barcelona-10

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239 - Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137 - Madrid-30
Tel. 430 84 81
Cataluña:
Lourdes Padrós
Manuel Girona, 61, 2.° - Barcelona-34
Teléfono 204 45 83

Controlado por O.J.D.

PROCEDENCIA DE LAS ILUSTRACIONES

Pintura de la portada de George V. Kelvin

Página	Fuente
12-13	Ilil Arbel
14-15	U.S. Army
16-17	Sovfoto
21	Laboratorio de Circuitos Integrados de la Universidad de Stanford, cortesía de Phillip W. Barth
22-26	Gabor Kiss
27	Laboratorio de Circuitos Integrados de la Universidad de Stanford, cortesía de Rosemarie Koch
28-29	Gabor Kiss
30	Laboratorio de Circuitos Integrados de la Universidad de Stanford, cortesía de Phillip W. Barth (*arriba*); Gabor Kiss (*abajo*)
31	Laboratorio de Circuitos Integrados de la Universidad de Stanford, cortesía de Phillip W. Barth
32	Gabor Kiss
33	Laboratorio de Investigación de IBM en San José, cortesía de Kurt E. Petersen
34-47	Jerome Kuhl
53-55	Instituto Oceanográfico de Woods Hole
56-65	Ian Worpole
69-70	Tom Prentiss
71	Tom Prentiss (*izquierda*), Adrian R. Morrison (*derecha*)
72-73	Adrian R. Morrison
74-76	Tom Prentiss
79	Ryszard Grygiel
80	Ryszard Grygiel (*arriba*), Peter Bogucki (*abajo*)
81-87	Alan D. Iselin
88-90	Patricia J. Wynne
91	Kenneth B. Raper
92	Patricia J. Wynne (*arriba*), John Tyler Bonner (*abajo*)
93	Patricia J. Wynne
94	Kenneth B. Raper
97	Ricardo Génova, F. García Olmedo y P. Carbonero Zalduegui
98	F. García Olmedo y P. Carbonero Zalduegui
99-100	Ricardo Génova, F. García Olmedo y P. Carbonero Zalduegui
101	M. A. Rodríguez-Loperena
102-103	Ricardo Génova, F. García Olmedo y P. Carbonero Zalduegui
108-111	Edward Bell
117	Matthew G. Wheeler
118-120	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Pedro IV, 160 - Barcelona-5
Fotocromos reproducidos por GINSA, S.A.
Imprime GRAFESA
Gráfica Elzeviriana, S. A.
Napoles, 249 - Tel. 207 40 11
Barcelona-13
Printed in Spain - Impreso en España

Cartas

Señor Director:

Después de leer la sección "Temas metamágicos", correspondiente al número del mes de abril, dedicada al LISP, quisiera hacerle los siguientes comentarios a la misma:

El traductor, que hizo un buen trabajo, debió dejar la sintaxis de las expresiones tal y como vienen en inglés, explicando, como lo hace en la página 114 con el átomo EQ, su significado en español. De hecho, si uno intenta darle a uno de los intérpretes LISP existentes una instrucción exactamente igual que la primera del texto: (SUMA 2 2) después de pulsar la tecla de retorno en vez de imprimir la cifra "4", nos imprimirá un mensaje anunciándonos que a la expresión le falta algo. Para ser más exactos, en el LISP que tenemos a mano, el mensaje exacto que devuelve es: "NO VALUE IS BOUND TO SUMA"; es decir, ningún valor está ligado a suma.

Prolépticamente hablando, el hecho de que LISP permita definir funciones no resuelve el problema; si haciendo uso de las facilidades del LISP, definiéramos previamente la función SUMA, tendríamos que hacerlo del modo siguiente o similar: (CSETQ SUMA (LAMBDA (A1 A2) (PLUS A1 A2))). Lo que además de ser una definición en principio puramente temporal, pues sólo dura el tiempo que dure el programa, la capacidad de escribirla implica un buen conocimiento de LISP. En consecuencia, hubiera uno podido ahorrarse todo el problema dándole directamente al LISP, en vez de (SUMA 2 2), (PLUS 2 2).

Por otra parte, no puedo menos de discrepar de la afirmación del autor de que el procedimiento interactivo sea más barato que el procedimiento de compilación. Desde un punto de vista individual, el uso de un procedimiento interactivo da más facilidades que cualquier otro, pero a expensas de una disminución de la eficiencia del sistema global. Por poner una analogía, es lo que le sucede a los organismos vivos respecto a la entropía: son capaces de disminuir su propia entropía a costa de aumentar la del sistema global o entorno.

En lo que concierne a la afirmación del autor, de que una vez que uno se acostumbra a la profusión de paréntesis las expresiones LIPS "se hacen intuitivas e incluso gratas de ver...", me parece cuando menos excesiva. Lo cierto es que aunque la sintaxis del LISP es sencilla y consistente, es difícil de leer. Algunos irónicamente afirman que lo que realmente quieren decir las siglas LISP es: "*L*ots of *I*nsipid *S*tupid *P*arenthesis"; es decir, montones de paréntesis insípidos y estúpidos. En cualquier caso, uno se encuentra, como dice el autor, "expresiones donde hay que terminar cerrando una docena de paréntesis", añado yo físicamente, sólo si quiere, al menos para algunos LISP. En efecto, en ciertos "dialectos" LISP el cerrar todos los paréntesis finales se puede hacer usando un símbolo concreto que para el caso del LISP instrumentado en el Centro de Proceso de Datos del Ministerio de Educación y Ciencia es el >.

Finalmente, estimo que el artículo, además de ser menos interesante que otros del mismo autor, resulta ocioso. Al que ya sabía LISP, no le aporta nada nuevo, mientras que, el primerizo, dudo mucho que leyéndose este artículo y los sucesivos con que nos "amenaza" el autor aprenda a programar en LISP, o incluso capte la idea de lo que significa el LISP en Inteligencia Artificial.

Esto no es óbice para que yo comparta las opiniones del autor sobre las cualidades y facilidades del LISP a las que yo le añadiría dos que considero básicas. La primera es la de encargarse de todo el "trabajo sucio" de programación, como es la reserva de memoria, almacenamiento de datos intermedios, etc. La segunda concierne a la superación de la dicotomía datos-instrucciones. La arquitectura de los ordenadores de tipo Von Neuman permitía que datos e instrucciones compartieran la memoria central del computador al banalizarlas. El LISP da un paso más al no distinguir datos de instrucciones. Esto tal vez sea una de las razones fundamentales por las cuales una de las características logicales "software" de la siguiente generación de computadores sea el uso del LISP como lenguaje básico.

JUAN PAZOS SIERRA

Prof. Inteligencia Artificial
Facultad de Informática
Universidad Politécnica de Madrid

Los autores

BARRY M. BLECHMAN y MARK R. MOORE ("Una zona desnuclearizada en Europa") son miembros del Centro Roosevelt de Estudios sobre Política Americana, de Washington. Blechman es vicepresidente del mismo y responsable del área de política exterior y política de defensa. Estudió en el Queens College de la Universidad de Nueva York, licenciándose en 1964. Se doctoró en 1971, por la Universidad de Georgetown. De 1966 a 1971 perteneció al Centro de Análisis Navales, instituto subvencionado por el Gobierno Federal de los Estados Unidos. Fue nombrado director adjunto de la Oficina Estadounidense de Control de Armamentos y Desarme, en 1977. Abandonó el Gobierno en 1979 para entrar en la Fundación Carnegie, y de donde pasó al Centro Roosevelt. Blechman participó en calidad de experto en la redacción del informe final de la Comisión Internacional para el Desarme y los Problemas de Seguridad, en el que presentaba el plan de creación de una zona desnuclearizada en Europa. Moore es ayudante de investigación del Centro Roosevelt. Se licenció en 1981 por el Amherst College.

JAMES B. ANGELL, STEPHEN C. TERRY y PHILLIP W. BARTH ("Dispositivos micromecánicos de silicio") son ingenieros eléctricos; realizaron el trabajo que sirve de base a este artículo cuando coincidieron profesionalmente en la Universidad de Stanford, donde Angell enseña ingeniería eléctrica y es secretario adjunto del departamento de ingeniería eléctrica. Cursó estudios en el Instituto de Tecnología de Massachusetts, licenciándose en 1946 y doctorándose en 1952. Melómano entusiasta, Angell tiene a su cargo el carillón de 35 campanas de la torre Hoover. Terry fundó Microsensor Technology, Inc. Es doctor en ingeniería eléctrica por Stanford, donde investigó hasta que fundara su compañía. Barth es investigador adscrito al laboratorio de circuitos integrados de Stanford. Se licenció por la Universidad de Notre Dame, en 1973, doctorándose en ingeniería eléctrica por Stanford, en 1980.

HAIM HARARI ("Estructura de quarks y leptones") ocupa la cátedra Annenberg de física de altas energías del Instituto Weizmann de Ciencias. De nacionalidad israelita obtuvo su licenciatura y su doctorado en física en la Universidad Hebrea de Jerusalén. Después de servir cuatro años en el ejército se incorporó, en 1966, en el Instituto Weizmann. Preside el comité de planificación y becas del Consejo de Enseñanza Superior, la institución responsable de administrar la ayuda financiera gubernamental para las universidades de Israel.

JOHN M. EDMOND y KAREN VON DAMM ("Fuentes termales en el fondo del océano") se han especializado en geoquímica marina. Edmond enseña esa disciplina en el Instituto de Tecnología de Massachusetts (MIT). Resume así su biografía: "Nací en Glasgow, en cuya universidad cursé la licenciatura en química. Posteriormente me trasladé a la Institución Scripps de Oceanografía, dependiente de la Universidad de California en San Diego, donde me doctoré con una tesis sobre el dióxido de carbono en agua. Lo que me abrió las puertas del claustro docente del MIT. En mi departamento seguimos tres líneas de investigación: geoquímica de los metales traza en el agua de mar y geoquímica marina general; química de las fuentes termales en la cresta de las dorsales y procesos químicos en grandes lagos y ríos y en sus estuarios. Entre mis aficiones se cuentan la política, historia y cultura de los países de ultramar donde tengo la oportunidad de realizar trabajos de campo". Von Damm realiza estudios de postgraduado en el Programa Coordinado de Oceanografía que llevan a cabo el MIT y la Institución Oceanográfica Woods Hole.

ADRIAN R. MORRISON ("Una ventana abierta al cerebro dormido") es profesor de anatomía de la Facultad de Veterinaria de la Universidad de Pennsylvania. Cursó los estudios de licenciatura en la Universidad de Cornell, doctorándose en anatomía (1964) por la de Pennsylvania. Después de un año de trabajo postdoctoral en la Universidad de Pisa volvió a Pennsylvania, donde ha permanecido desde entonces.

PETER BOGUCKI y RYSZARD GRYGIEL ("Primeros agricultores de la llanura del Norte de Europa") vienen trabajando, desde 1976, en el estudio de las culturas de los primeros asentamientos agrícolas del centro y norte de Polonia. Bogucki es catedrático de antropología de la Universidad de Massachusetts en Boston. Se ocupa, sobre todo, del análisis de los huesos animales desenterrados en los yacimientos de las comunidades agrícolas. Grygiel dirige el departamento de Neolítico del Museo Arqueológico y Etnográfico de Łódź. Se formó en la Universidad Adam Mickiewicz, de Poznań; en 1980 se recibió de doctor en arqueología por la de Varsovia. Ha profundizado en la cronología de las primeras comunidades neolíticas del centro y norte de Polonia, amén de excavar muchos yacimientos pertenecientes al primer Neolítico.

JOHN TYLER BONNER ("Señales químicas de agregación en las amebas") enseña biología en la cátedra George M. Moffett y es jefe del departamento de biología de la Universidad de Princeton. Procede de la Universidad de Harvard. "En 1939, cuando cursaba todavía la licenciatura, empecé a interesarme por los hongos mucosos (o amebas sociales); veía en ellos unos organismos especialmente aptos para desentrañar la biología del desarrollo. Se trató sin duda de un flechazo, pues desde entonces hemos avanzado en compañía. Licenciado ya, y después de servir en las Fuerzas Aéreas, me trasladé a Princeton, donde continúo. Además de proseguir mi trabajo experimental, he escrito varios libros, la mayoría en torno al tema general de la evolución y el desarrollo."

FRANCISCO GARCIA OLMEDO y PILAR CARBONERO ZALDUEGUI ("El control genético de las proteínas del trigo") han venido realizando durante los últimos años investigaciones sobre distintos aspectos de la biología molecular de las plantas cultivadas, en el departamento de bioquímica de la Escuela Técnica Superior de Ingenieros Agrónomos de la Universidad Politécnica de Madrid. García Olmedo, director de dicho departamento desde 1970, es licenciado en ciencias químicas y doctor ingeniero agrónomo. Ha sido investigador posdoctoral y, posteriormente, profesor visitante en la Universidad de Minnesota. Pertenece a diversas sociedades científicas, habiendo sido vicepresidente de la Sociedad Europea para la Investigación en Mejora de Plantas (EUCARPIA) y de la Sociedad Española de Bioquímica (SEB). Carbonero Zaldueguí es doctor ingeniero agrónomo y profesor agregado de bioquímica. Ha sido becario de la Fundación Juan March en el Departamento de Ciencia y Tecnología de Alimentos en la Universidad de Minnesota y, posteriormente, investigadora del Instituto Nacional de Investigaciones Agrarias.

Hace...

José M.ª López Piñero

... doscientos años

Alexandre-Gui Pingré incluyó en su *Cométographie*, traducido al francés, el informe que José de Zaragoza había remitido a la Academia de Ciencias, de París, acerca de sus observaciones del cometa de 1664.

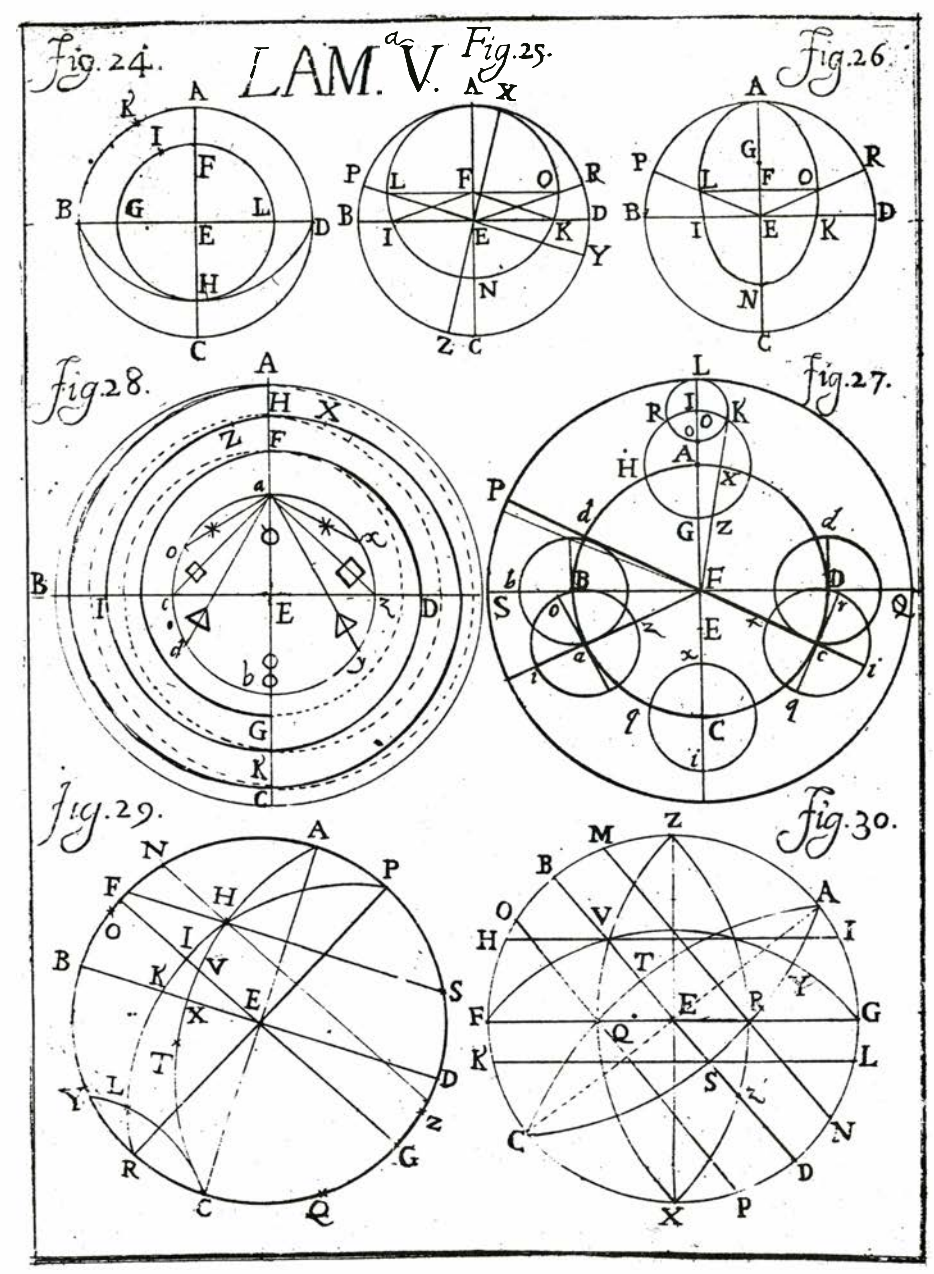
José de Zaragoza y Vilanova (1627-1679) nació en la localidad castellonense de Alcalá de Chisvert y estudió en la Universidad de Valencia. Inclinado desde muy joven a las matemáticas, le ofrecieron una de las dos cátedras de esta disciplina existentes en dicha institución, que no llegó a aceptar porque ingresó a los veinticuatro años en la Compañía de Jesús. Tras residir en distintas ciudades españolas, fue nombrado profesor en el colegio que los jesuitas tenían en Mallorca. Allí estuvo en relación con varios matemáticos y astrónomos, recibiendo en especial el influjo de Vicente Mut, que fue su auténtico maestro. Tras enseñar una breve temporada en los colegios de Barcelona y Zaragoza fue trasladado, con anterioridad a 1660, al de San Pablo de Valencia, ciudad en la que residió más de un decenio, dedicado casi exclusivamente al estudio y a la enseñanza de las matemáticas y a la observación astronómica. Allí editó sus primeras obras y fue maestro de algunos científicos y técnicos, como José Vicente del Olmo y José Chafrión. A finales de 1670 fue nombrado titular de la cátedra de matemáticas del Colegio de San Isidro de Madrid, donde permaneció los nueve años que le restaban de vida, desempeñando otros cargos oficiales como consejero científico y como maestro de matemáticas del monarca. Publicó entonces la mayor parte de sus trabajos.

Las obras matemáticas de Zaragoza responden fundamentalmente a una intención didáctica. La primera que imprimió fue la titulada *Aritmética universal* (1669), compendio elemental de aritmética y álgebra. Para poder publicarla tuvo que hacer personalmente los caracteres tipográficos propios del álgebra, que faltaban en las imprentas españolas: "Apliqué mi industria y conseguí lo que sólo intentar pareció a muchos temeridad. Hice por mi mano los punzones, matrices y llaves; fundí todos los caracteres enteros y quebrados que juzgué necesarios, sin perdonar trabajo ni gasto, por conseguir toda perfección". Tanto este texto como la *Geometría especulativa y práctica* (1671) y sus posteriores ediciones latinas, significaron una notable elevación del nivel dentro del empobrecido panorama español de la disciplina, pero no introdujeron ningún elemento innovador. La *Geometría* en concreto estaba basada en la traducción de varios libros de Euclides con nueva presentación pedagógica.

Mayor interés tiene su libro titulado *Trigonometría española*, cuya primera redacción terminó en Valencia el año 1663, pero que no fue impresa hasta 1672. Incluye las primeras tablas de logaritmos publicadas en España y consiste en una síntesis de diferentes obras anteriores, entre las que se encuentra la del español Caramuel. A pesar de ello, en su edición latina de 1673, Zaragoza criticó acertadamente el sistema de logaritmos "perfectos" ideado por este último, demostrando que no eran otra cosa que los complementos de los logaritmos vulgares. No obstante, su mejor producción matemática fue la *Geometria magna in minimis* (1674), primer texto español que incluye teoremas geométricos distintos de los de Euclides, y que incluso tiene algunos rasgos aislados de originalidad, si bien desconectados de la evolución general europea.

La importancia a escala nacional de estos textos de Zaragoza y el mérito

1. Los sistemas astronómicos, según la **Esphera en común celeste y terráquea** *(1675), de José de Zaragoza*



2. *Esquemas relativos a las observaciones y cálculos astronómicos de José de Zaragoza. La figura 26 corresponde a la primera ley de Kepler, que Zaragoza expone así: "Keplero y Bulialdo quieren que los planetas no formen círculo con su movimiento natural sino una figura oval, que se llama* **elipse**, *como AINK; y el punto F que está en medio del diámetro AN es centro de la* **elipse** *y tiene dos puntos E. G. que distan igualmente del centro F, de calidad que la suma de las dos rectas que salgan de G E a cualquiera punto de la circunferencia sea siempre igual al diámetro mayor AN, y estos dos puntos GE se llaman* **focos** *o* **fuegos** *de la elipse. Los planetas, pues, en esta sentencia forman una* **elipse** *que la Tierra E es uno de sus focos. Y porque el centro de la* **elipse** *F no es el centro del mundo E, será la* **elipse** *excéntrica y el apogeo del planeta es A y el perigeo N, como en el círculo antecedente. Esta sentencia es singular y especiosa y no produce malos efectos, en el cálculo de los planetas, como veremos en la Astronomía".*

personal de su esfuerzo no deben hacer olvidar los defectos de su información, que pesaron negativamente en una eficaz renovación de la matemática española. Peñalver, por ejemplo, anotó la ausencia en sus escritos de los transcendentales progresos que el álgebra debía a Cardano, Vieta y Descartes. Este autor, sin embargo, subraya su juicio matemático y sus condiciones de expositor serio y concienzudo. "Por más que su fama no alcanza el mismo grado que la de Caramuel –afirma– fue, a mi juicio, un matemático superior a éste."

En contraste con las limitaciones de su obra matemática, la labor astronómica de Zaragoza fue auténticamente valiosa y estuvo conectada con las nuevas corrientes todo lo ampliamente que las circunstancias españolas permitían. Fue, por de pronto, un gran astrónomo práctico. Desde 1660, al menos, durante su segunda estancia en Valencia, y quizá con anterioridad, a partir de su relación en Mallorca con Vicente Mut, sabemos que hacía observaciones con anteojos. Estas observaciones culminaron en las de los cometas de 1664 y 1677. El informe de la correspondiente al primero, como hemos dicho, fue incluido por Pingré en su *Cométographie*. Zaragoza se refiere a dicho cometa en los siguientes términos: "El año 1664 a medio Diciembre apareció un cometa de los más célebres por su duración, extendido curso y variedad, pues pasó de matutino a vespertino, de austral a boreal, de retrógrado a directo, y se pudo observar de todo el mundo; su cola hizo un semicírculo desde el Poniente por Septentrión al Levante. Observáronle en Italia Juan Cassini; en Francia, el Padre Ignacio Pardies; en Lima, D. Francisco Ruiz de Lozano; en Mallorca, Don Vicente Mut, Astrónomo insigne, y sacó a luz sus observaciones en nada inferiores a las de Tycho; yo lo observé en Valencia desde 17 de Diciembre de 1664 hasta 21 de Marzo de 1675. Comencé a observarle en 4 gr. de Libra con latitud austral 24.19 bajo retrógrado hasta 49 gr. de latitud austral en 29 gr. de Cancro, luego volvió a subir hasta la eclíptica y la cortó en 29 gr. de Ariete, prosiguió retrógrado siempre hasta los 26.44 de Ariete, con 5 gr. 49 min. de latitud boreal, y a 7 de Febrero se hizo directo y hasta los 21 de Marzo volvió al primer grado de Tauro, subiendo hasta 9 gr. de latitud boreal. Corrió en su curso retrógrado 160 gr. de longitud".

Su observación del cometa de 1677 fue la primera del mundo y sus noticias fueron publicadas en el *Journal des Savants* y en las memorias de la Academia de Ciencias, de París. Zaragoza fue un experto constructor de instrumentos científicos e incluso ideó varios astronómicos y otros de aplicación geométrica y topográfica, que expuso en una monografía de la que más tarde nos ocuparemos. Entre los instrumentos que fabricó se encontraban los dos anteojos astronómicos con los que realizó la mayor parte de sus observaciones.

A este intenso cultivo de la astronomía práctica unía Zaragoza una excelente información de los datos de la ciencia astronómica de su tiempo. Ello le permitió publicar su importante obra titulada *Esphera en común, celeste y terráquea* (1675). De acuerdo con su título, comprende tres libros. El primero, "De la esphera en común", es de carác-

ter geométrico y trata de las líneas y círculos que pueden trazarse en ella. El segundo, "De la esphera celeste", es el consagrado en sentido estricto a la astronomía. El tercero, "De la esphera terrestre", es la primera exposición sistemática publicada en España de lo que hoy llamamos geofísica.

La obra está apoyada en un conocimiento de primera mano de los principales autores, tanto antiguos y medievales como modernos. Copérnico y Tycho Brahe, Galileo, Kepler, Descartes, Gassendi, Cassini y también científicos jesuitas como Clavius, Fabri y Kircher, se encuentran entre los más citados. No hay que pensar, sin embargo, que se trata de un mero resumen didáctico. Su ideología es claramente moderna, ante todo por la continua fundamentación de las hipótesis en datos de observación astronómica, en buena parte propios. En segundo lugar, por la explícita aplicación de los nuevos datos a la crítica de las bases de cosmología clásica. Por ejemplo, Zaragoza niega la incorruptibilidad de la sustancia celeste, admite la infinitud del espacio, se opone al alma de los cielos, así como a los orbes cristalinos o sólidos de Aristóteles y Santo Tomás. La única limitación de importancia es su actitud en el problema de los sistemas planetarios, cuestión en la que mantiene una extremada cautela que evita toda adhesión terminante. Esta cautela sólo puede explicarse por su condición de partidario en secreto del heliocentrismo y del movimiento de la Tierra.

En el detenido análisis que Cotarelo dedicó a la obra astronómica de Zaragoza se aducen abundantes pruebas de la adhesión del astrónomo valenciano a la doctrina de Copérnico. Al tratar de los diversos sistemas astronómicos la expone la última, alterando el orden cronológico, y la califica de "ingeniosa aunque condenada". Su posición es que "por modo de hipótesis o suposición pueden todos valerse de ella para el cálculo de los planetas; con que sólo se condena la actual realidad de esta composición pero no su posibilidad", rebuscada distinción que le sirve para justificar el hecho de que el sistema de Copérnico sea la base implícita de todas las teorías astronómicas parciales que defiende. Del mismo modo, al analizar las apariencias de la inmovilidad de la Tierra, afirma que, a pesar de su evidencia sensorial, "se puede responder que se engaña el sentido, porque la grandeza de los cielos es tanta, que la distancia de la Tierra y centro es como un punto respecto de la grandeza de los cielos, y así la vista no puede observar diferencia en las estrellas...".

Una zona desnuclearizada en Europa

La prohibición de todo armamento nuclear en una franja de terreno que se extienda a ambos lados de la frontera entre el Este y el Oeste, parece una medida políticamente factible para reducir el riesgo de guerra atómica

Barry M. Blechman y Mark R. Moore

Las conversaciones que vienen desarrollándose en Ginebra entre los Estados Unidos y la Unión Soviética para el control del armamento nuclear apuntan directamente hacia dos objetivos: reducir el número de armas nucleares estratégicas, o de largo alcance, desplegadas por las dos superpotencias y reducir el número de las de alcance medio, capaces de destruir objetivos en Europa. Queda excluida de las negociaciones otra importante categoría de armas nucleares: las tácticas, o de corto alcance. De las aproximadamente 6000 armas nucleares actualmente desplegadas en Europa por las fuerzas militares de la Organización del Tratado del Atlántico Norte (OTAN), más de la mitad están clasificadas como tácticas. Se considera que las fuerzas del Pacto de Varsovia poseen un arsenal semejante al otro lado de la frontera centroeuropea. Las armas nucleares de corto alcance desplegadas por ambas partes en Europa tienen como misión proporcionar apoyo directo a las fuerzas terrestres que entren en combate, por cuya razón suelen denominarse también armas nucleares de teatro. En tal epígrafe se incluyen las ojivas nucleares para misiles balísticos de corto alcance (con un alcance máximo de 150 kilómetros), obuses de artillería (con un alcance máximo de 30 kilómetros), misiles antiaéreos y minas explosivas.

El elevado número de armas nucleares de teatro instaladas en Europa presenta un problema especial para el control de armamentos. Dado que muchas se hallan desplegadas en posiciones avanzadas, próximas a los frentes de cualquier eventual guerra, tienen grandes posibilidades de verse arrolladas por las fuerzas invasoras. Ante una disyuntiva de hacer uso de las mismas o perderlas, el bando defensor probablemente se vería presionado a utilizarlas,

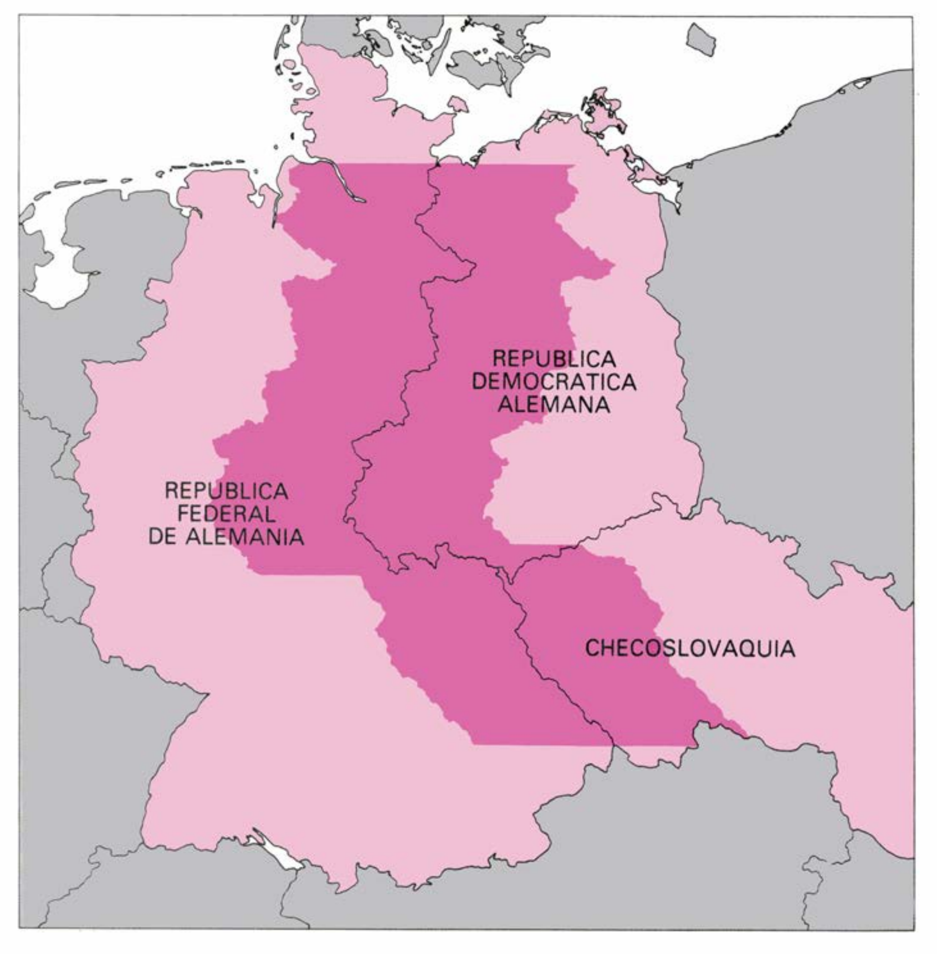
NOTA DEL EDITOR

La Comisión Independiente para el Desarme y los Problemas de Seguridad, que lanzó un llamamiento en favor de la creación de una zona desnuclearizada en su último informe, hecho público en junio de 1982, estaba formada por 17 miembros, todos los cuales lo eran a título personal. La presidía Olof Palme, diputado entonces al Parlamento Sueco, presidente del Partido Socialdemócrata Sueco y antiguo primer ministro, puesto que ocupa de nuevo. Los otros miembros eran Giorgi Arbatov, miembro del Comité Central del Partido Comunista de la Unión Soviética, diputado al Soviet Supremo y director del Instituto de los Estados Unidos y Canadá, una división de la Academia de Ciencias; Egon Bahr, diputado al Parlamento de la República Federal de Alemania y presidente de su subcomité para el desarme y el control de armamentos, y antiguo ministro de cooperación económica; Gro Harlem Brundtland, diputado al Parlamento Noruego, secretario del Partido Laborista Noruego y antiguo primer ministro; Jozef Cyrankiewicz, antiguo primer ministro de Polonia y presidente que fue del Consejo de Estado; Jean-Marie Daillet, diputado al Parlamento Francés, vicepresidenté de su comité de defensa y presidente del comité de defensa del partido giscardiano UDF; Robert A. D. Ford, consejero especial del gobierno canadiense para las relaciones Este-Oeste y antiguo embajador de Canadá en Colombia, Yugoslavia, Egipto y la Unión Soviética; Alfonso García-Robles, embajador-presidente de la delegación de México en el Comité de las Naciones Unidas para el Desarme, antiguo primer ministro de México y premio Nobel de la Paz (1982) por su papel en la creación de una zona desnuclearizada en Latinoamérica; Haruki Mori, antiguo embajador de Japón ante S. M. Británica y ante la Organización para la Cooperación y el Desarrollo Económico (OCDE) y antiguo viceministro de asuntos exteriores; C. B. Muthamma, embajador de la India en los Países Bajos, y anteriormente en Ghana y Hungría; Olusegun Obasanjo, general, miembro del Consejo de Estado, miembro del Consejo de la Universidad de Ibadán y antiguo Jefe de Estado de Nigeria; David Owen, diputado al Parlamento Británico y antiguo secretario de Estado para asuntos exteriores y de la Commonwealth; Shridath Ramphal, secretario general de la Commonwealth y antiguo ministro de asuntos exteriores de Guyana; Salim Salim, ministro de asuntos exteriores de Tanzania; Soedjatmoko, rector de la Universidad de las Naciones Unidas en Tokio y antiguo embajador de Indonesia en Estados Unidos; Joop den Uyl, diputado al Parlamento y vice-primer ministro de los Países Bajos, líder del Partido Laborista Holandés y antiguo primer ministro; Cyrus Vance, antiguo secretario de Estado de los Estados Unidos.

extendiéndose, por tanto, el conflicto más allá del "umbral nuclear".

En la profunda crisis política que posiblemente precedería a la ruptura de las hostilidades, los responsables políticos sufrirían una considerable presión para delegar de antemano en los comandantes militares destacados en Europa la autoridad de hacer uso de las armas nucleares de teatro en determinadas circunstancias. Es dudoso que un presidente norteamericano o un gobernante europeo llegasen a tomar tal decisión, pero es seguro que se ejercería una fuerte presión en ese sentido, por lo que no debe descartarse por completo tal posibilidad. Aunque no se llegara a tomar deliberadamente esa decisión, el mantenimiento eficaz del control de las armas nucleares de teatro se haría difícil en tiempos de guerra. Dada la confusión y la desinformación que inevitablemente acompañan a las operaciones militares de importancia, la ubicación concreta de los emplazamientos de las armas nucleares sería un factor de crucial relieve para determinar si se toma o no la decisión de iniciar una guerra nuclear.

Para abordar el problema, el pasado año presentó un nuevo plan la Comisión Independiente sobre Desarme y Problemas de Seguridad (Independent Commission on Disarmament and Security Issues), grupo internacional compuesto por personalidades políticas actualmente en el poder o que lo estuvieron en otros tiempos, creado en 1980 y presidido por Olof Palme, entonces líder de la oposición y hoy nuevamente primer ministro de Suecia. La Comisión, cuyos miembros procedían de países del Este, del Oeste y del Tercer Mundo, contaba con personalidades como Cyrus Vance, ex secretario de Estado de los Estados Unidos y Giorgi Arbatov, miembro del Comité Central del Partido Comunista de la Unión Soviética. En un informe publicado en junio de 1982, la Comisión lanzaba un llamamiento para el establecimiento, en Europa, de una "zona libre de armas nucleares de teatro", junto con otras "medidas para reforzar el umbral nuclear y reducir las presiones para el uso anticipado de armas nucleares".



1. ZONA DESNUCLEARIZADA propuesta para Europa central, de la que se retirarían todas las armas nucleares. Se reduciría así el riesgo de que, en caso de un ataque militar convencional, el bando a la defensiva se viera forzado a ser el primero en hacer uso de sus armas nucleares de teatro, en un intento de evitar que las fuerzas invasoras lo arrollaran. Según el informe de la Comisión Palme, la determinación geográfica precisa de la zona "debería determinarse en negociaciones" entre las partes implicadas. A efectos ilustrativos, la comisión sugirió, no obstante, una zona que abarcaría en un principio una franja de 150 kilómetros de anchura a ambos lados de la frontera de la República Federal de Alemania con la República Democrática Alemana y Checoslovaquia (*color oscuro*). A diferencia de anteriores propuestas de este tipo, podría continuar estacionándose armas nucleares en parte de los tres estados (*color claro*).

Según el informe de la Comisión, *Seguridad común: anteproyecto para la supervivencia*, la zona desnuclearizada propuesta comenzaría en Europa central, tal vez con una franja de 150 kilómetros de anchura a cada lado de la frontera que divide la República Federal de Alemania de la República Democrática Alemana y Checoslovaquia. Con el tiempo, la zona iría ampliándose hasta abarcar los flancos septentrionales y meridionales de ambas alianzas. Dentro de la zona quedaría prohibido todo tipo de armamento nuclear.

La idea de una zona europea desnuclearizada no es nueva. Pero la propuesta de la Comisión se diferencia de todas las anteriores en varios aspectos significativos, no siendo el menor el hecho de que se base en las características de una categoría específica de armas nucleares, en lugar de cualquier otro criterio, como por ejemplo la extensión territorial de los estados implicados. Desde nuestro punto de vista, el plan de la Comisión constituye una vía práctica y políticamente factible para reducir de inmediato el riesgo de una guerra nuclear en Europa. En este artículo resumiremos los argumentos en favor de una zona centroeuropea libre de armas nucleares de teatro; en particular, explicaremos por qué el plan podría tener éxito allí donde todos los demás fracasaron.

La mayoría de los estados coinciden en mantener libres de armas nucleares ciertas zonas: la Antártida, el espacio exterior y los cuerpos celestes, los fondos marinos e Iberoamérica están sujetos a tratados que prohíben el despliegue de armas nucleares en su terreno. También ha habido propuestas en el sentido de prohibir las armas nucleares en el Oriente Medio, Asia meridional, el Pacífico sur y Africa, así como en Europa. Todas y cada una de estas proposiciones han sido objeto de discusión en diversos foros internacionales. Pero la historia de los esfuerzos para crear zonas desnuclearizadas ha producido más deseos que resultados, y las propuestas parecen a menudo maniobras diplomáticas dirigidas a obtener ventajas políticas, y no intentos genuinos de reducir el riesgo de guerra nuclear.

El primer y más minucioso plan en favor de una zona desnuclearizada en Europa fue presentado en 1957 por Adam Rapacki, a la sazón ministro de asuntos exteriores de Polonia. El plan

Rapacki, como se le conoció, postulaba que los gobiernos de Polonia, Checoslovaquia, República Democrática Alemana y República Federal de Alemania llegasen al acuerdo de no fabricar, no mantener y no permitir la instalación de armas nucleares dentro de sus fronteras. El status desnuclearizado de la zona quedaría garantizado por los estados que entonces disponían de la bomba atómica (Estados Unidos, Unión Soviética y Gran Bretaña) y se implantaría un sistema de vigilancia aérea y terrestre para asegurar su cumplimiento. En su propuesta original, el plan Rapacki no se presentaba como un tratado formal que delimitara la zona, sino que descansaba en declaraciones unilaterales de los gobiernos interesados.

Los objetivos políticos del plan Rapacki parecían evidentes a las naciones occidentales. Se había presentado justo cuando los gobiernos de los países de la OTAN debatían la posibilidad de compensar la superioridad militar convencional (no nuclear) de las fuerzas del Pacto de Varsovia mediante un "escudo nuclear". Ciertamente, poco después, en la reunión del Consejo del Atlántico Norte de diciembre de 1957, se decidió adoptar tal política. Resultado de ello fue que los países de la OTAN consideraron el plan Rapacki un intento de debilitar su posición militar en Europa central; y lo rechazaron.

El plan polaco se revisó y sometió a la consideración de las naciones occidentales dos veces, en noviembre de 1958 y en marzo de 1962. Las modificaciones introducidas no dieron satisfacción a las principales objeciones de la OTAN y apenas se avanzó. Las posteriores propuestas del Este en favor de la designación de zonas desnuclearizadas europeas se concentraron en subregiones del continente. Los Balcanes y la región del Adriático fueron durante mucho tiempo objeto de propuestas, al igual que toda la cuenca mediterránea. La región nórdica, que el primer ministro de la Unión Soviética Nikita Jrushov sugirió ya en 1959 como una zona a desnuclearizar, ha vuelto a ser en los últimos años centro de gran atención por parte de las autoridades soviéticas.

Poco después de que se iniciara la presentación de propuestas de creación de zonas desnuclearizadas por parte del Este, los Estados Unidos elaboraron una lista de los cuatro requisitos que debería cumplir toda iniciativa para que recibiera el apoyo norteamericano: (1) la propuesta partiría voluntariamente de los estados de la región; (2) incluiría a todos los estados de la región, o al menos a aquellos que contaran con un poder militar importante; (3) no deberían alterarse el equilibrio militar ni los acuerdos en materia de seguridad; (4) el acuerdo recogería cláusulas de comprobación de su cumplimiento.

	ARMA	PRIMER DES-PLIEGUE	ALCANCE (KILO-METROS)	POTENCIA MAXIMA (KILO-TONES)	NUMERO DE RAMPAS DE LAN-ZAMIENTO
OTAN	LANCE	1973	115	50	90
	HONEST JOHN	1953	38	20	90
	MORTERO M109 DE 155 MILIMETROS	1964	18	2	1455
	MORTERO M110 DE 203 MILIMETROS	1962	16	1	390
	PLUTON (FRANCIA)	1974	120	10	42
PACTO DE VARSOVIA	FROG/SS-21	1950-60/ 1977	75/155	200	375
	SCUD-A	1957	85	1	250
	MORTERO DE 203 MILIMETROS	1979	31	?	150
	MORTERO DE 240 MILIMETROS	1979	30+	?	150

2. ARMAS NUCLEARES DE TEATRO actualmente desplegadas a ambos lados de la frontera centroeuropea. Comprende esa categoría los misiles balísticos de corto alcance y las piezas de artillería que se contabilizan en esta tabla. En la relación de las fuerzas de la OTAN se incluyen todos los sistemas de armas asignados al teatro europeo de operaciones. (Francia es miembro de la OTAN, aunque no está integrada en la estructura militar de la Alianza.) En la relación de las fuerzas del Pacto de Varsovia se incluyen todos los sistemas estacionados fuera de la Unión Soviética y en las regiones militares occidentales de la misma. Todos los misiles reseñados en la tabla pueden transportar solamente una ojiva nuclear cada uno. No se incluyen las minas explosivas nucleares ni los misiles nucleares tierra-aire, por clasificarse a veces tanto éstos como aquéllas como armas nucleares tácticas. También se excluyen los miles de armas nucleares de largo alcance desplegadas en tierra o mar que pueden alcanzar objetivos europeos.

En el caso de Iberoamérica, los Estados Unidos consideraron, como es obvio, que la propuesta de crear una zona desnuclearizada cumplía aquellos requisitos, o podía modificarse para que los cumpliera. Por el contrario, los Estados Unidos y otros gobiernos occidentales no han considerado aptas para la negociación las anteriores propuestas para la creación de zonas desnuclearizadas en Europa. Algunas objeciones han sido de carácter técnico: dada la constante oposición de la Unión Soviética a la inspección sobre el terreno, parece difícil, si no imposible, asegurar la retirada y permanente ausencia de armas nucleares en una región que las posee en número muy elevado.

No obstante, las principales objeciones de los occidentales a la creación de zonas desnuclearizadas en Europa han tenido su razón de ser en otros factores: la simple asimetría de las características geopolíticas de los Estados Unidos y de la Unión Soviética y el papel fundamental que las armas nucleares desempeñan en la estrategia militar de Occidente y en las relaciones políticas entre los miembros de la OTAN. Por el peso de estas razones, las autoridades occidentales se han inclinado a considerar las propuestas orientales de creación de zonas desnuclearizadas, en diversas partes de Europa, altamente propagandísticas y dirigidas a provocar la disensión entre los países de Europa Occidental y la ruptura de los lazos políticos entre los miembros de la alianza. Hay que comprender las reservas de la OTAN si se quiere establecer un plan de medidas realistas para reducir el riesgo de una guerra nuclear en el viejo continente.

Desde la creación de la alianza atlántica, las autoridades y los ciudadanos de los países occidentales han venido discutiendo hasta qué punto es eficaz y prudente apoyarse en el armamento nuclear para mantener la paz en Europa. Los abogados de la defensa convencional argumentan que el camino más seguro para lograr la estabilidad en Europa es mantener fuerzas militares convencionales de contundencia suficiente para persuadir a los dirigentes soviéticos de que, en caso de conflicto, serían derrotados en el campo de batalla y que, por tanto, se verían obligados a escoger entre la escalada de la guerra nuclear o aceptar la derrota. Si los dirigentes soviéticos creyesen que semejante dilema sería el posible resultado de un conflicto en Europa, siguen argumentando, se decidirían en contra de la guerra, optando por favorecer las relaciones estables entre los países del Este y el Oeste. La defensa de esa postura ha llevado a realizar constantes esfuerzos para potenciar las fuerzas convencionales de la OTAN.

Para muchos occidentales la amenaza de una guerra nuclear y las fuerzas necesarias para hacer verosímil tal

amenaza constituyen la razón de ser de la OTAN. Desde este punto de vista, lo que incita a los dirigentes de la Unión Soviética a buscar la estabilidad y evitar la guerra en Europa no es el cálculo de una posible derrota, sino la impresión de que, en el caso de que los acontecimientos se les escapasen de las manos, los resultados del enfrentamiento militar podrían ser desastrosos. De hecho, quienes mantienen este punto de vista argumentan que las fuerzas militares de la OTAN deberían estar pensadas para evitar una situación en la que los resultados de un conflicto pudieran calcularse. Si puede preverse el resultado, sostienen, éste podría ser aceptable en ciertas condiciones. Vale más, prosiguen en su argumentación, mejorar la disuasión inculcando la total convicción de que el resultado inexorable de un conflicto constituiría una catástrofe general.

Quienes mantienen este punto de vista pretenden mostrar a los dirigentes soviéticos que cualquier acto de manifiesta agresión en Europa podría fácilmente degenerar en un conflicto y, finalmente, convertirse en una guerra nuclear total entre los Estados Unidos y la Unión Soviética. En resumen, se inclinan por una posición en la que el peligro último de confrontación en Europa, el holocausto nuclear, se haga bien patente. Y ninguna mejora de las fuerzas convencionales puede servir a tal fin. Desde este punto de vista, cuanto más fuertes parezcan las defensas convencionales, mayor será el riesgo de que los dirigentes soviéticos yerren en sus cálculos y lleven a cabo una agresión, en la creencia de que tales defensas han reducido el riesgo de una guerra nuclear.

Por esa razón, entre otras, la OTAN nunca ha cumplido plenamente el objetivo acordado de fortalecer su dotación convencional. También por ello declara explícitamente la OTAN que, si fuese necesario, desencadenaría la escalada de una guerra nuclear en Europa. Igualmente es esa la razón de que la OTAN mantenga miles de armas tácticas en Europa Occidental, esté desplegando una nueva generación de armas nucleares de alcance medio y exprese periódicamente su preocupación por el equilibrio entre las fuerzas nucleares estratégicas de Estados Unidos y la Unión Soviética.

Hay un motivo más para insistir en la defensa nuclear. Desde hace casi 40 años, los estrategas y los políticos norteamericanos sostienen que la Unión Soviética representa la amenaza más peligrosa para los intereses de su país. A la Unión Soviética se la considera no sólo un peligro para la seguridad norteamericana en términos militares, sino también una amenaza, a través de medios más sutiles, para los valores de la sociedad occidental. De ahí que, si analizamos la estrategia norteamericana en términos geopolíticos, vemos que su hilo conductor ha consistido, a lo largo de este período, en la contención del poder militar y la influencia política de la Unión Soviética en la masa continental de Eurasia. Más importante aún: la estrategia de los Estados Unidos se ha dirigido a limitar la influencia de la Unión Soviética en las regiones de Europa situadas más allá del cinturón de países ocupados por los ejércitos soviéticos en 1945. El vehículo escogido para tal estrategia ha sido, obviamente, la OTAN.

En la lucha entre las dos superpotencias por la influencia en Europa Occidental, los Estados Unidos se hallan en desventaja geográfica. El poder militar de la Unión Soviética constituye una presencia inmutable que se siente en las capitales de Europa Occidental. Frente a ello, y a pesar de la gran apuesta política y económica de los Estados Unidos en la continua independencia y en la orientación proamericana de Europa Occidental, el compromiso de seguridad realizado por los Estados Unidos y su presencia militar en Europa son vistos, y siempre lo serán, como algo transitorio e incierto. Los europeos son conscientes de que los Estados Unidos tienen siempre la alternativa de retirarse al otro lado del océano que les resguardó en el pasado. Tales temores nunca podrán disiparse del todo. Antes bien, se agravan cada vez que los Estados Unidos sufren uno de sus recurrentes ataques de neoaislacionismo.

Para compensar la asimetría, tanto percibida como real, de las presencias militares norteamericana y soviética en Europa, los responsables políticos estadounidenses vienen realizando continuos esfuerzos para robustecer los lazos entre las fuerzas americanas estacionadas en Europa y el conjunto del poder militar de los Estados Unidos, y para tranquilizar a sus aliados –y prevenir a sus adversarios– acerca de la duración de sus compromisos. Tales esfuerzos adoptan formas diversas, pero su principal contenido lo constituyen los pasos dados para "conectar" las fuerzas estratégicas nucleares a las fuerzas terrestres de los Estados Unidos en Europa.

La estrategia occidental en Europa se basa en un supuesto de respuestas escalonadas a la agresión, comenzando con contrataques convencionales, si el ataque del Pacto de Varsovia se limitase a fuerzas convencionales, pero incluyendo la posibilidad de iniciación y escalada de guerra nuclear. Esta postura, bien patente, cuenta con el peso del



3. MISIL NORTEAMERICANO LANCE. Aparece en los preparativos para una prueba de lanzamiento desde su vehículo-rampa autopropulsado. Este misil de combustible líquido transporta una sola carga, que puede ser convencional o estratégica. Su alcance máximo es de 150 kilómetros. El Lance ha sustituido al Honest John como misil nuclear convencional de teatro. Otros países de la OTAN tienen en sus arsenales ambos misiles. Todas las ojivas nucleares de los misiles se hallan bajo custodia estadounidense.



4. MORTERO NORTEAMERICANO M110 DE 203 MILIMETROS. Aparece aquí realizando una prueba de tiro con un obús no nuclear; puede también lanzar cargas nucleares. Su alcance máximo es de unos 16 kilómetros. Puede ir montado sobre un vehículo autopropulsado, como se ve en la fotografía, o sobre chasis de remolque. La estructura trasera absorbe el retroceso. Los Estados Unidos han suministrado M110 a la mayoría de sus aliados europeos.

despliegue de armas nucleares de corto alcance en Europa, no lejos de los presuntos frentes. Tales armas, que se han proyectado para su empleo contra las fuerzas militares en el teatro de operaciones, podrían ser las primeras en utilizarse en muchos escenarios, si se tomase la decisión de traspasar el umbral nuclear. Por tener una potencia bastante pequeña y apuntar contra objetivos militares, se supone que harían más verosímil la amenaza de iniciación de una guerra nuclear.

Para aumentar la credibilidad de su amenaza nuclear, la OTAN proyecta desplegar en Europa misiles estadounidenses de alcance medio capaces de bombardear objetivos situados en territorio soviético. Si el empleo de las armas nucleares de teatro no pusiera fin al conflicto, podrían lanzarse armas de alcance medio, siempre sobre objetivos militares. Esa escalada comportaría menos dificultades que esta otra: la intervención de las fuerzas estratégicas centrales de los Estados Unidos. Aún así, el impacto nuclear sobre territorio soviético por parte de misiles norteamericanos controlados conduciría, casi con seguridad, a una respuesta rusa contra los Estados Unidos, lo que precipitaría la intervención de las fuerzas estratégicas centrales de ambos bandos. Para entendernos: puesto que los dirigentes de la Unión Soviética reconocerían esa escalada, ello les disuadiría de dar el primer paso. En teoría se refuerza la disuasión hasta el punto de dotar de verosimilitud a una intervención de las fuerzas nucleares estratégicas norteamericanas.

En cierto modo, ese reparto del riesgo es el vínculo último, desde el punto de vista práctico y psicológico, que une a la alianza atlántica, al menos en lo que respecta a las cuestiones de seguridad. Con su promesa de recurrir, si fuese necesario, a las fuerzas nucleares estratégicas en defensa de Europa Occidental, y aceptando, en consecuencia, la eventualidad de una represalia soviética sobre su territorio, los Estados Unidos se ligan simbólicamente al destino de Europa, compensando así su aislamiento geográfico natural.

A la luz de este razonamiento, se aprecian claramente, desde el prisma occidental, los defectos de las propuestas del Este para la desnuclearización de ciertas regiones de Europa. Una zona desnuclearizada, en su propuesta tradicional, privaría a Europa Occidental de un vínculo de vital importancia entre sus defensas convencionales y la disuasión estratégica nuclear de los Estados Unidos, acentuando así las asimetrías existentes, tanto entre las situaciones geográficas de ambas partes, como en el equilibrio de fuerzas militares convencionales. Este razonamiento se aplica en primer lugar a las propuestas (como el plan Rapacki) de creación de zonas desnuclearizadas en Europa central, subregión que se halla en el corazón de todas las reflexiones sobre la defensa de Europa Occidental. Pero también se dan algunos de esos defectos políticos y militares en propuestas parecidas presentadas para otras subregiones; en todo caso, se las considera precedentes nada útiles que podrían hacer más difícil el rechazo de la creación de una zona desnuclearizada en el corazón del viejo continente.

Desde el punto de vista de la OTAN, la debilidad de las zonas desnuclearizadas, tal como se ha planteado hasta ahora, puede expresarse en los siguientes términos. En primer lugar, las iniciativas no se han visto generalmente acompañadas de propuestas de restricción de fuerzas convencionales. Ello implica que se vería disminuida la capacidad de la OTAN para compensar la mayor potencia de las fuerzas convencionales del Pacto de Varsovia con la amenaza de una escalada nuclear. Naturalmente, la OTAN podría movilizar sus recursos económicos, más poderosos, para competir con mayor eficacia en el terreno militar convencional, lo que sería costoso y parece improbable, vista la tradicionalmente escasa predisposición de las democracias occidentales a soportar pesadas cargas militares en tiempos de paz.

Además, aunque se instrumentasen defensas convencionales más eficaces, quedaría en el aire el interrogante de hasta qué punto una disminución de la amenaza nuclear de la OTAN reduciría la capacidad de los países europeos para resistir presiones políticas por parte

de la Unión Soviética, o incluso a una agresión militar. Aunque difícilmente podría probarse, el hecho de que las relaciones internacionales en Europa hayan evolucionado hacia una configuración estable en los últimos 35 años sugiere que la estrategia atlántica tal vez haya dado buen cumplimiento a las expectativas militares y políticas de sus creadores. Lo que viene a indicar que la doctrina actual no debería modificarse sin una razón poderosa que permitiese esperar que las alternativas tengan el mismo éxito y que sean más adecuadas a las presentes circunstancias.

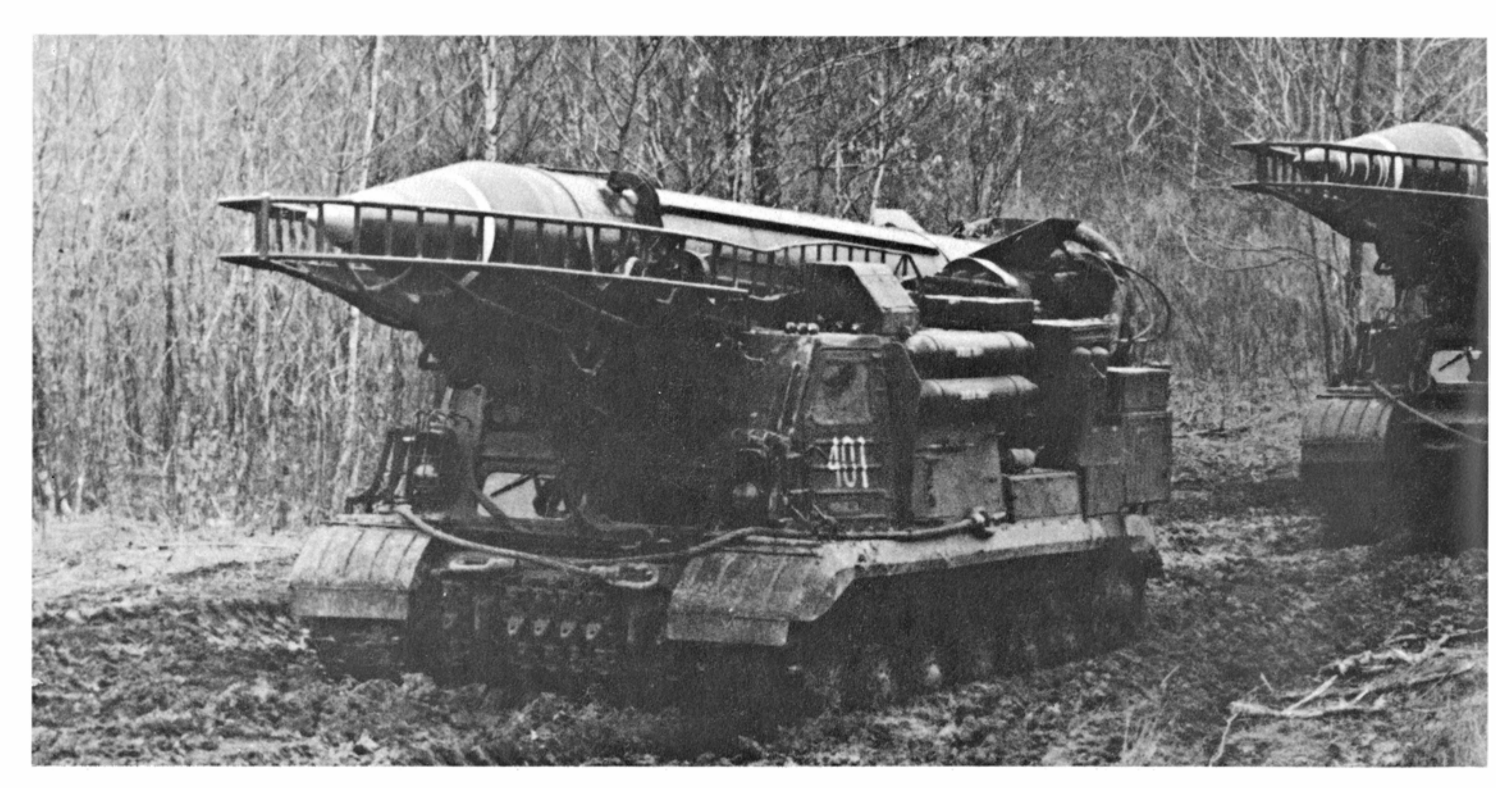
En segundo lugar, dado que la retirada de fuerzas nucleares de Europa central implicaría cierto desenganche de la disuasión estratégica nuclear norteamericana de la defensa de Europa Occidental, podría verse en la misma un presagio de separación de los Estados Unidos de Europa Occidental. Tal eventualidad, se teme, conduciría inevitablemente a un aumento de la influencia soviética en los asuntos europeos, e incluso a un posterior control de los mismos. Es de por sí esclarecedor que no se incluyera el territorio soviético en las propuestas para desnuclearizar Europa central. Significa en realidad que, al aceptar la OTAN el plan Rapacki o cualquiera de los posteriormente propuestos, se admitiría una disminución de la presencia norteamericana en Europa Occidental, sin una disminución equivalente de la presencia soviética, más poderosa. Las implicaciones políticas de tan desventajosa negociación saltan a la vista. Se podría interpretar como una cesión de influencia decisoria a la Unión Soviética en los asuntos de Europa Occidental.

Se contempla, además, otra cuestión en los cálculos de la OTAN. La incorporación formal de la República Federal de Alemania a una alianza militar formada por naciones de Europa Occidental y de América del Norte tenía como propósito, además de la contención del poderío soviético, el de restañar las heridas de la segunda guerra mundial. La OTAN ha promovido, juntamente con otras organizaciones económicas internacionales creadas en los años 50, la reconciliación entre Alemania y sus antiguos enemigos, permitiendo el desarrollo de relaciones estables en Europa e impidiendo las tensiones nacionales que condujeron a tres desastrosas guerras en un solo siglo.

El plan Rapacki situaba a la República Democrática y a la República Federal alemanas, juntamente con Polonia y Checoslovaquia, en una categoría especial. La República Federal habría dejado de ser miembro de pleno derecho de la alianza. Además, semejante decisión no habría sido tomada por los propios alemanes, sino que les habría sido impuesta por presiones exteriores. Ello habría acentuado la especial situación germana, aumentando las posibilidades de aparición de fuerzas nacionalistas que pudiesen poner en peligro sus lazos políticos con los demás estados de Europa Occidental, e incluso los lazos existentes entre el Este y el Oeste. En resumen, por dar un tratamiento especial a Alemania, el plan Rapacki y las demás propuestas de zonas centroeuropeas desnuclearizadas despertaron el temor de ver reavivarse viejas tensiones. Más aún, en opinión de muchos políticos occidentales, debería proseguir la integración de Alemania con los restantes países de Europa Occidental y la mejora gradual de las relaciones entre las dos alianzas.

Así, el plan Rapacki y las posteriores propuestas para el establecimiento de zonas desnuclearizadas en Europa central se vio frustrada, sobre todo, por sus implicaciones de índole política. Al sugerir una clara diferenciación de riesgos entre los Estados Unidos y Europa, se interpretaron como un medio para deteriorar la influencia norteamericana en Europa e incrementar el poderío soviético. Por considerar aparte a Alemania, dándole un especial tratamiento, se creyó que podrían socavar los esfuerzos llevados a cabo para integrar Alemania Federal en la comunidad occidental de naciones. La actual propuesta, presentada por la Comisión Independiente, en el sentido de establecer una zona desnuclearizada, ha tenido buen cuidado de evitar esos inconvenientes.

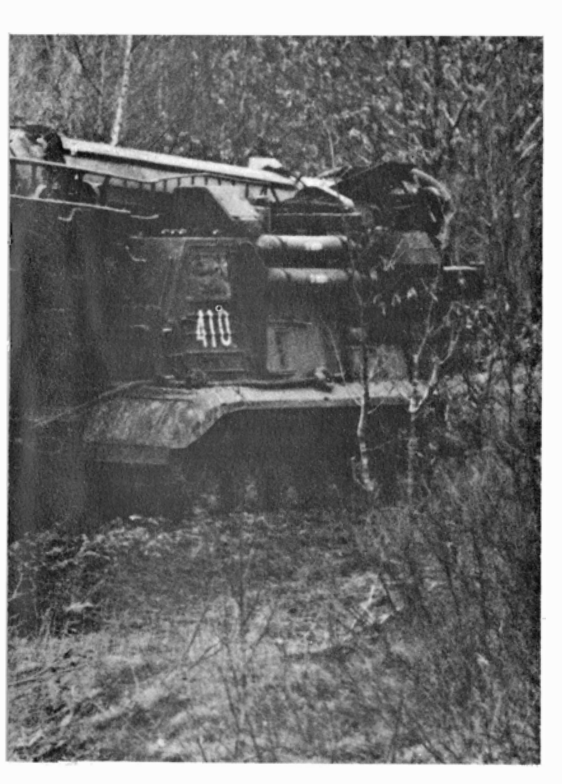
La comisión tomó buena nota de la completa realidad de las relaciones entre los diferentes tipos de fuerzas



5. MISILES SOVIETICOS DE CORTO ALCANCE transportados en vehículos-rampas de lanzamiento erguibles y sobre cadenas. Estos misiles, denominados SCUD-A por los servicios de información occidentales, transportan una ojiva, que puede ser tanto nuclear como convencional. Su alcance

militares instaladas en Europa e instó a adoptar una perspectiva global en el tema del control de armamentos como único camino para llegar a niveles más bajos de fuerzas y reducir el riesgo de guerra. Declaró en su informe que "un acuerdo negociado de cuasi-paridad de fuerzas convencionales entre las dos alianzas... facilitaría la reducción de armas nucleares y la alteración de la prioridad que hoy se otorga a los armamentos nucleares en la planificación de la contingencia nuclear".

En consecuencia, la comisión comenzó sus recomendaciones para Europa instando a la convocatoria de una conferencia de los ministros de asuntos exteriores de los países de la OTAN y del Pacto de Varsovia, que han venido llevando a cabo negociaciones en Viena durante los últimos diez años para un acuerdo sobre fuerzas convencionales. Se requería a los ministros de asuntos exteriores a que llegasen a un preacuerdo que garantizase la paridad entre ambas partes y que posibilitase la continuación de las negociaciones para una reducción substancial de las fuerzas convencionales. En opinión de los miembros de la comisión, las diferencias existentes entre ambas partes podrían resolverse con bastante prontitud de existir la voluntad política de superarlas. La comisión concluía afirmando que el acuerdo de celebrar una reunión de ministros de asuntos exteriores proporcionaría el impulso necesario y una oportunidad para seguir avanzando.



máximo es de unos 85 kilómetros. Las ojivas nucleares del Este se encuentran bajo custodia soviética.

La comisión prestó detenida atención a las armas químicas, instando su retirada de Europa y proponiendo medidas complementarias para asegurar que ambas partes observaran las prohibiciones. Estas últimas medidas incluirían inspecciones recíprocas sobre el terreno; es decir, que cada parte, si así lo deseaba, podría inspeccionar anualmente un determinado número de instalaciones, elegidas por ella, en el territorio de la otra. Es esta una propuesta importante e imaginativa. El hecho de que el representante soviético en la comisión estuviese de acuerdo en ello permite aventurar una mejor acogida que en el pasado, por parte soviética, a las exigencias occidentales de cuidadosa atención a los problemas de comprobación de los acuerdos de control de armamentos. La acogida soviética a las preocupaciones de comprobación se hizo también patente en la discusión de la zona desnuclearizada propuesta.

La comisión examinó, desde diferentes ángulos y actitudes alternativas, el componente nuclear de los problemas de la seguridad europea, incluyendo la creación de zonas desnuclearizadas definidas por fronteras nacionales. Al final, sin embargo, los miembros de la comisión decidieron seguir una vía funcional, concentrándose en categorías específicas de armas que plantean problemas concretos.

La comisión prestó especial atención a los "problemas planteados por aquellas armas nucleares cuyos sistemas de lanzamiento se encuentran desplegados en número considerable en posiciones europeas avanzadas". Instó a la retirada mutua de las armas nucleares de teatro de una zona desnuclearizada. Para reducir las dificultades de comprobación, la comisión propuso que no se permitiese en la zona ninguna munición nuclear. Establecía, además, que deberían prohibirse también los silos de almacenamiento de municiones nucleares, las maniobras de simulación de operaciones nucleares y los preparativos para la colocación de minas nucleares. Finalmente, se apuntaba la necesidad de negociar una regulación especial para el despliegue en la zona de artillería de "doble finalidad" y de misiles de corto alcance, que podrían usarse para transportar munición nuclear.

La comisión consideró mejor no recomendar fronteras concretas para la zona, dejando tal cometido a las partes implicadas. A efectos ilustrativos, el informe indicaba, sin embargo, la posibilidad de que se señalara una zona de 300 kilómetros centrada en la frontera Este-Oeste [*véase la figura 1*]. Dicha zona incluiría en principio parte de la República Federal y de la República Democrática alemanas y de Checoslovaquia. A diferencia del plan Rapacki y de las posteriores propuestas del Este, quedaría todavía territorio de los tres estados disponible para el despliegue de armas nucleares. Esta era la diferencia más importante. Conforme la zona se extendiese de norte a sur, según se estableciesen sus límites, parte de Austria, Hungría, Yugoslavia, Italia, Dinamarca, Suecia, Noruega, Finlandia y la Unión Soviética podrían verse incluídas en ella.

La comisión dejó también a los negociadores la definición concreta de las medidas necesarias para garantizar a las partes la posibilidad de comprobar el cumplimiento de lo pactado. Sin embargo, los miembros acordaron que tales medidas "incluyeran un número limitado de inspecciones recíprocas sobre el terreno". Es de destacar, una vez más, que el representante soviético estuviese de acuerdo en que se incluyera esta fase. Aunque Arbatov, como todos los demás miembros de la Comisión, lo era a título personal, es evidente que no daba su asentimiento al texto, y especialmente a las recomendaciones, sin una autorización superior. En este contexto, la aceptación de las inspecciones sobre el terreno está en consonancia con la tendencia hacia una mayor aceptación por parte soviética de medios de comprobación más intrusionistas.

Finalmente, la comisión recomendaba medidas adicionales para reforzar el umbral nuclear y reducir las presiones para el uso anticipado de las armas nucleares. La recomendación con mayor carga de intención fue que no se desplegasen en Europa aquellas armas nucleares que presentan mayores visos de empleo, especialmente las pequeñas armas nucleares de radiación intensa, como la "bomba de neutrones". Aunque no se hacía ninguna recomendación en concreto, una parte del informe discutía también los problemas específicos de los misiles antiaéreos y de las minas explosivas de carga nuclear, sugiriendo que su retirada de Europa podría reforzar los obstáculos al desencadenamiento de una guerra nuclear.

En resumen, la zona desnuclearizada propuesta por la Comisión parece no presentar las imperfecciones de los planes anteriores. ¿Qué se conseguiría con su implantación?

Por un lado, se reforzarían los obstáculos contra el desencadenamiento imprevisto o accidental de una guerra nuclear. Parece clara la improbabilidad de que una guerra nuclear resulte de la fría

apreciación de una posición ventajosa en tiempos de normalidad, pues el poderío de ambas partes hace insignificante cualquier ventaja prevista. Es más probable que la guerra nuclear surja de decisiones desesperadas tomadas en una grave crisis, probablemente en una crisis en la que ya estén implicadas las fuerzas convencionales de ambas partes. En tales circunstancias, cualquiera de ellas podría iniciar un ataque nuclear, al no ver otra alternativa para conjurar una derrota catastrófica, o por haber llegado a la conclusión de que el adversario está a punto de llevar la escalada al nivel nuclear.

La retirada de armas nucleares de una ancha zona a ambos lados de la frontera centroeuropea ofrecería más posibilidades de reducir la eventualidad de su ocurrencia que de precipitar tales decisiones desesperadas. Las armas estarían fuera del caos de la guerra, al menos en los primeros momentos. Se daría a los líderes políticos de ambas partes el bien más precioso: tiempo. Permitiría un análisis más cuidadoso y profundo de la situación militar táctica y la posibilidad de continuar en el uso exclusivo de las fuerzas convencionales. Reduciría la preocupación de ambas partes de ver forzada a la otra a deponer rápidamente las armas y, consiguientemente, disminuiría la premura de lanzar un ataque preventivo. Finalmente, permitiría una mayor exploración de las alternativas diplomáticas: siendo mayor el tiempo disponible antes de decidir el lanzamiento de armas nucleares, una y otra parte podrían llegar a tomar conciencia de los extraordinarios peligros de la carrera en la que se hallasen embarcadas y darían los pasos necesarios para resolver el conflicto antes de que se les fuese de las manos.

Llevar a cabo la propuesta de la comisión de una zona desnuclearizada ofrecería otras ventajas. Los cambios necesarios para llegar a tal acuerdo podrían llevar por separado a ambas partes a reconsiderar los inventarios actuales de armas nucleares instaladas en el continente. Muchos observadores han sostenido la viabilidad de importantes reducciones en lo militar, pero han señalado su preocupación por las implicaciones políticas de una retirada unilateral de armamentos. Los reajustes derivados del establecimiento de una zona desnuclearizada, y la convicción de que la otra parte habría de llevar a cabo cambios semejantes, permitirían una reducción importante de las fuerzas nucleares sin sufrir las consecuencias políticas adversas.

La negociación fructífera de una zona desnuclearizada, unida a las medidas propuestas para restringir las armas convencionales y químicas, abriría posibilidades aún más ambiciosas de controlar y reducir los arsenales militares en Europa y en otras regiones. Tendría efectos positivos en las relaciones políticas entre el Este y el Oeste y daría paso a unas relaciones menos tensas y de mayor cooperación entre los Estados Unidos y la Unión Soviética. Sería también positiva para las relaciones en el seno de la alianza atlántica, eliminando lo que ha constituido una perjudicial causa de fricción entre los Estados Unidos y Europa Occidental: el debate sobre diferentes apreciaciones del control de armas nucleares en el continente.

A finales de 1982, el gobierno sueco solicitó formalmente a todos los gobiernos de Europa y de América del Norte que expresasen sus puntos de vista sobre la zona desnuclearizada propuesta. La Unión Soviética respondió en enero declarando que se sumaba a la idea y que plantearía la propuesta en las conversaciones de Viena. También señaló, no obstante, que la actual propuesta no iba muy lejos, añadiendo que la zona debería alcanzar entre 500 y 600 kilómetros de anchura. Tal implicación comprendería la práctica totalidad de la República Federal y la República Democrática alemanas.

Hasta ahora los países occidentales no han mostrado ningún entusiasmo ante la idea de una zona desnuclearizada. Pero si los soviéticos presentan realmente la propuesta en Viena, sería una oportunidad inmejorable para ligar la creación de una zona desnuclearizada a la conclusión de un pre-acuerdo sobre limitación de fuerzas convencionales y para negociar una anchura de la zona mutuamente aceptable que no planteara problemas políticos como los del plan Rapacki.

La creación de una zona desnuclearizada no eliminaría de raíz el riesgo de una guerra nuclear en Europa. Lo habrá mientras los armamentos nucleares permanezcan en el inventario de las grandes potencias. El peligro de guerra aumenta o disminuye en consonancia con las relaciones políticas y los cambios en los despliegues militares de ambas partes, pero no es probable que desaparezca. Es esa una realidad con la que hay que vivir. Pero la zona desnuclearizada propuesta constituye una medida práctica y políticamente factible para controlar el peligro de guerra nuclear en Europa, que merece seria atención por parte de todos los ciudadanos y gobiernos.

Dispositivos micromecánicos de silicio

Válvulas, boquillas, espejos, sensores de presión y otros sistemas mecánicos se graban químicamente en un monocristal de silicio en forma de oblea. Todos estos dispositivos se fabrican ya en serie

James B. Angell, Stephen C. Terry y Phillip W. Barth

A lo largo de los últimos 30 años el silicio se ha convertido en la materia prima para la fabricación de componentes y sistemas electrónicos. Bajo la forma de elemento microelectrónico o pastilla ("chip"), el silicio se encuentra en sistemas tan diversos como las máquinas lavavajillas o los dispositivos de control de los transbordadores espaciales. La utilización de sus propiedades eléctricas se ha visto acompañada por una menos conocida exploración de otras características y aplicaciones potenciales suyas. Esta investigación ha conducido al desarrollo de una tecnología conocida con el nombre de micromecanización, que permite construir dispositivos mecánicos de silicio casi tan pequeños como los microelectrónicos.

La micromecanización se inicia con las mismas técnicas de fabricación en serie que han abaratado los elementos de circuitos integrados de silicio. Estas técnicas posibilitan la fabricación simultánea de un gran número de elementos; en consecuencia, el coste de producción se reparte entre todos ellos. La micromecanización se basa en estos mismos principios y utiliza además las técnicas de grabado químico; con todo ello es capaz de producir una amplia serie de elementos tridimensionales: pozos, agujeros, pirámides, surcos, hemisferios, puentes, diafragmas, agujas y muros. Por combinación de estos elementos estructurales se construyen múltiples dispositivos mecánicos.

Entre los micromecanismos de silicio que se han construido cabe citar válvulas, muelles, espejos, toberas, conectadores, cabezas impresoras, tableros de circuitos, sumideros de calor y sensores de magnitudes tales como fuerza, presión, aceleración y concentración química. Incluso un dispositivo de la complejidad del cromatógrafo de gases, instrumento que identifica y mide la proporción en que se encuentran unos gases en una mezcla desconocida, se puede construir en un disco de silicio de escasos centímetros de diámetro.

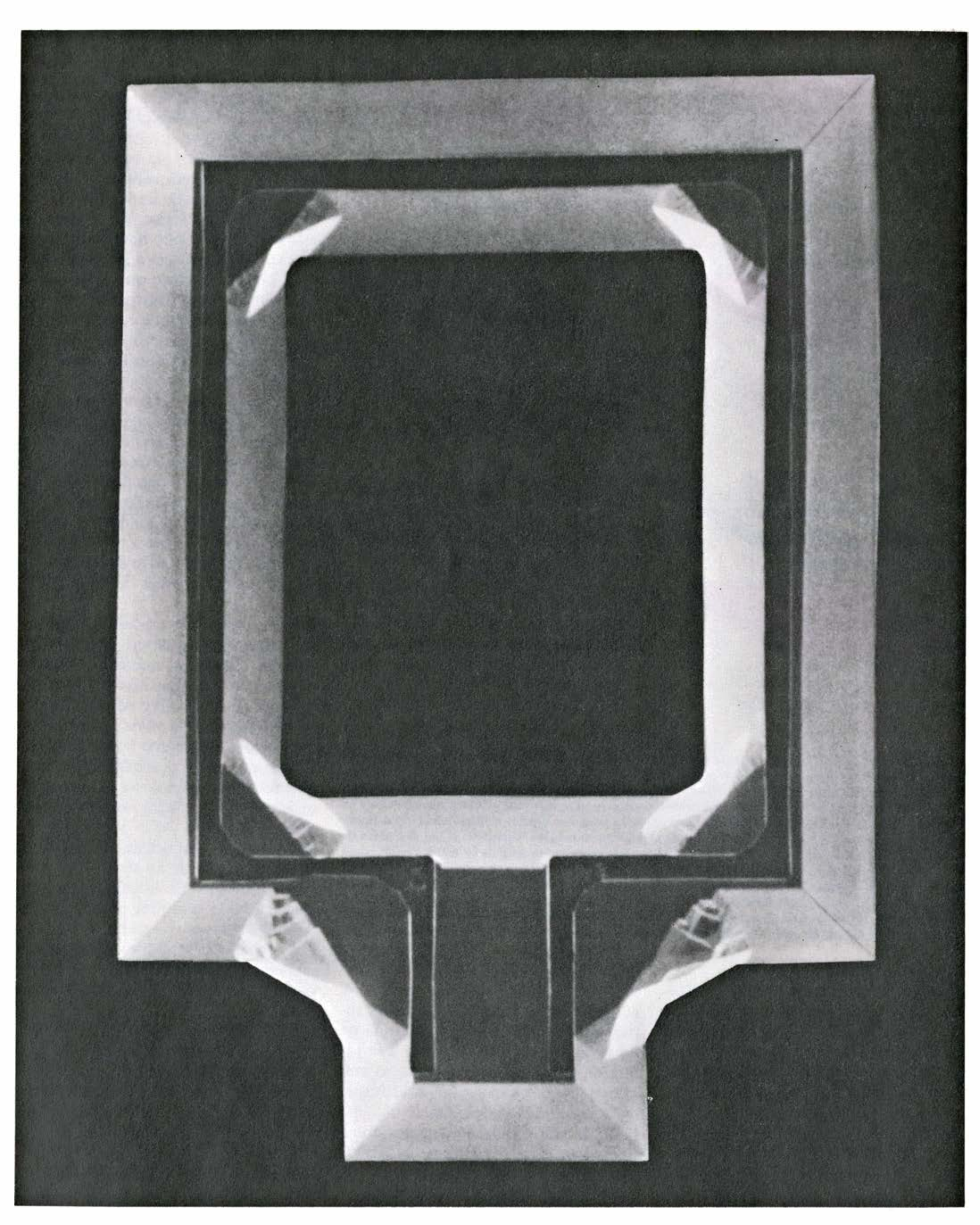
Gran parte del interés de los dispositivos micromecánicos radica en la necesidad de disponer de sensores que sean lo más económicos y versátiles posible. Hasta épocas recientes, los componentes electrónicos de los sistemas de control automáticos y de los instrumentos de medida costaban más que los sensores. Con la llegada de los microprocesadores, el coste de todos estos componentes electrónicos se abarató considerablemente. Los sensores fabricados en serie reemplazaron a los sensores montados manualmente, mucho más caros, de la misma forma que los circuitos de silicio fabricados en serie sustituyeron a los dispositivos de válvulas de vacío.

El último avance en este campo es el sensor integrado, un elemento de silicio que incluye tanto el sensor como la electrónica que procesa subsiguientemente la señal. Los sensores integrados son más económicos que los sensores ordinarios y los componentes electrónicos necesarios fabricados por separado. Por otra parte, debido a que la señal es menos vulnerable al ruido y a las pérdidas, los sensores integrados se comportan mejor que los sensores discretos y permiten una amplia variedad de aplicaciones.

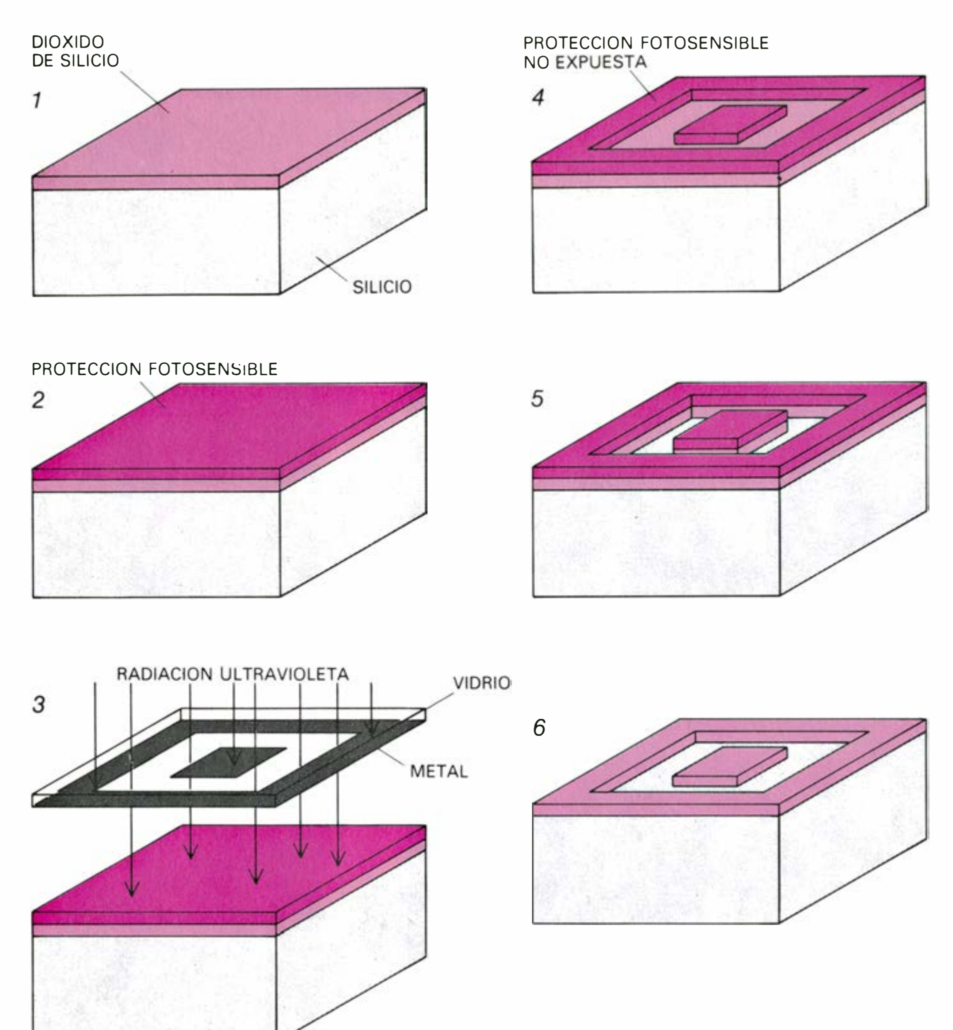
El silicio es un semiconductor, uno de los elementos situados entre los metales y los no metales en el sistema periódico de los elementos. Sin embargo, se distingue de los otros semiconductores porque se oxida con facilidad. El silicio forma una capa superficial de dióxido de silicio (SiO_2) cuando se expone a la acción del vapor de agua. Esta capa es químicamente inerte y un aislante eléctrico; esencialmente es un vidrio. Las capas de óxido se utilizan para proteger ciertas zonas del silicio durante el proceso de fabricación en serie de dispositivos microelectrónicos. Este es el motivo por el cual la mayoría de estos dispositivos están fabricados en silicio y no con otros semiconductores, germanio por ejemplo. La micromecanización hace amplio uso de estas capas superficiales de óxido de silicio.

El silicio que se emplea en la industria electrónica toma la forma de monocristales cuya producción es, en sí misma, una técnica muy refinada, basada en la preparación de monocristales de silicio en forma de barras de 10 centímetros de diámetro por un metro de longitud; estas barras se cortan posteriormente en rodajas finas de 0,2 a 0,5 milímetros de grosor. A continuación, se pulimenta esas obleas hasta conseguir un acabado especular. La estructura cristalina homogénea del material le proporciona las propiedades eléctricas que se necesitan en los circuitos microelectrónicos. Por otra parte, el silicio posee también excelentes propiedades mecánicas.

El silicio está situado inmediatamente debajo del carbono en el sistema periódico y da lugar al mismo tipo de cristal que el diamante, aunque los enlaces interatómicos son un tanto más débiles. Los monocristales de silicio, frágiles, quiebran como los diamantes, pero su dureza es mayor que la de la mayoría de los metales. Los monocristales de silicio se muestran sorprendentemente resistentes a los esfuerzos mecánicos. Tanto en la tensión como en la compresión, su límite de elasticidad es mayor que el del acero; por otra parte, cuando se alcanza dicho límite, el silicio se fractura en tanto que el acero se deforma inelásticamente. Por último, los mono-



1. ACELEROMETRO DE SILICIO desarrollado por Lynn M. Roylance, de la Universidad norteamericana de Stanford. Este instrumento miniaturizado aprovecha las propiedades mecánicas y eléctricas del silicio. Se fabrica mediante un ataque selectivo de una lámina en forma de oblea cortada de un monocristal de silicio de grandes dimensiones. La figura es una microfotografía electrónica de uno de estos acelerómetros. En líneas generales, el instrumento consta de una masa de silicio con un espesor igual al de la oblea original suspendida del extremo de una cinta delgada de silicio. En la parte posterior de dicha cinta se ha colocado un resistor eléctrico, cuya resistencia varía cuando la cinta se flexiona. Este fenómeno nos proporciona una medida de la aceleración. La masa principal del acelerómetro se separa del material que la rodea mediante un proceso de grabado químico. La cinta se forma atacando la totalidad del silicio base, salvo una pequeña capa delgada. El dispositivo al que nos estamos refiriendo tiene una superficie de dos por tres milímetros y su espesor es de 0,6 milímetros. Sus dimensiones son lo suficientemente pequeñas como para unirse por sutura a un corazón humano con el fin de medir su aceleración. Partiendo de una única oblea de silicio se pueden fabricar simultáneamente centenares de acelerómetros como el reproducido en esta figura.



2. FIGURAS DE OXIDO existentes en la superficie de una oblea de silicio con el fin de proteger ciertas áreas seleccionadas durante el proceso de ataque químico húmedo que se utiliza para producir un dispositivo micromecánico. En primer lugar, se forma una capa de óxido de silicio (SiO_2) sobre la superficie; el silicio reacciona con el vapor de agua a temperaturas comprendidas entre 800 y 1200 grados Celsius y da lugar al óxido. La oblea oxidada se recubre de un polímero protector sensible al ultravioleta. A continuación, la oblea se coloca bajo una plantilla maestra (una figura metálica depositada sobre una placa de vidrio) a través de la cual se expone a dicha radiación. La sustancia fotosensible se debilita en las zonas expuestas a la radiación y posteriormente es eliminada mediante un baño de revelado. Más adelante, la oblea se somete al ataque de un ácido grabador; este ácido actúa sobre el óxido no protegido, pero no sobre la sustancia fotosensible residual o el silicio. El resultado de todo el proceso es una figura que constituye una réplica exacta de la figura metálica depositada sobre la placa de vidrio. En las etapas subsiguientes la figura de óxido se utiliza como máscara para el grabado del silicio situado bajo ella.

cristales de silicio se mantienen inalterados bajo ciclos repetidos de tensiones y compresiones, en tanto que los metales policristalinos tienden a debilitarse y romperse debido a las tensiones acumuladas en las zonas intercristalinas.

Tanto la micromecanización como la fabricación microelectrónica se inicia con la fotolitografía, la técnica fotográfica utilizada para transferir copias, a partir de un original, a la superficie del disco de silicio. El primer paso de la fotolitografía consiste en hacer crecer una capa delgada de óxido sobre la superficie del disco calentándola entre 800 y 1200 grados Celsius en una atmósfera de vapor de agua. (También se puede utilizar oxígeno seco, pero la acción del vapor de agua es mucho más rápida.) A continuación, se deposita sobre la superficie de óxido una capa delgada de un polímero orgánico sensible a la radiación ultravioleta. En contacto con esta placa se coloca una plantilla especial, constituida generalmente por una placa de vidrio sobre la que existe una figura metálica; el conjunto se expone a la radiación ultravioleta. El metal de la pantalla es opaco a esta radiación; el vidrio, transparente. La radiación origina una reacción química en las zonas expuestas de la lámina. Si la sustancia protectora fotosensible es del tipo positivo, la reacción rebaja el polímero, mientras que si es del tipo negativo la reacción fortalece el polímero. La lámina se sumerge en un líquido revelador capaz de eliminar las zonas expuestas o las no expuestas, con lo que se obtiene sobre la superficie de la lámina una figura con óxido al descubierto y con óxido protegido por la capa fotosensible sobre la superficie de la lámina. Esta figura es la imagen positiva o negativa del original grabado en la plantilla.

Más tarde la lámina se sumerge en una solución de ácido fluorhídrico, que ataca al óxido pero no a la sustancia fotosensible ni al silicio que está situado bajo ella. La capa fotosensible protege las zonas del óxido que cubre. Una vez que el óxido expuesto ha sido eliminado, la sustancia fotosensible residual se puede eliminar a su vez con ácido sulfúrico caliente, dado que éste ataca a la sustancia fotosensible pero no al óxido ni al silicio. El resultado final de todo el proceso es una figura de óxido sobre la superficie del disco que duplica exactamente la imagen de la capa protectora fotosensible y que, por tanto, es una copia positiva o negativa de la plantilla. La figura de óxido sirve de plantilla en los pasos subsiguientes del proceso.

Es en este punto en donde la micromecanización difiere de la fabricación de elementos microelectrónicos. Al procesar un dispositivo electrónico, la figura de óxido sirve de plantilla durante la adición a la oblea de impurezas ("dopado") tales como boro o fósforo. Al calentar el disco, los átomos de impurezas depositados sobre su superficie se difunden ligeramente en el silicio no recubierto por el óxido y se crean así las regiones conductivas o resistivas que son la base de los dispositivos microelectrónicos. En la micromecanización, el óxido se utiliza como máscara durante el proceso de grabado químico. Los grabadores atacan el silicio no cubierto por la capa de óxido y excavan pozos tridimensionales profundos en el disco.

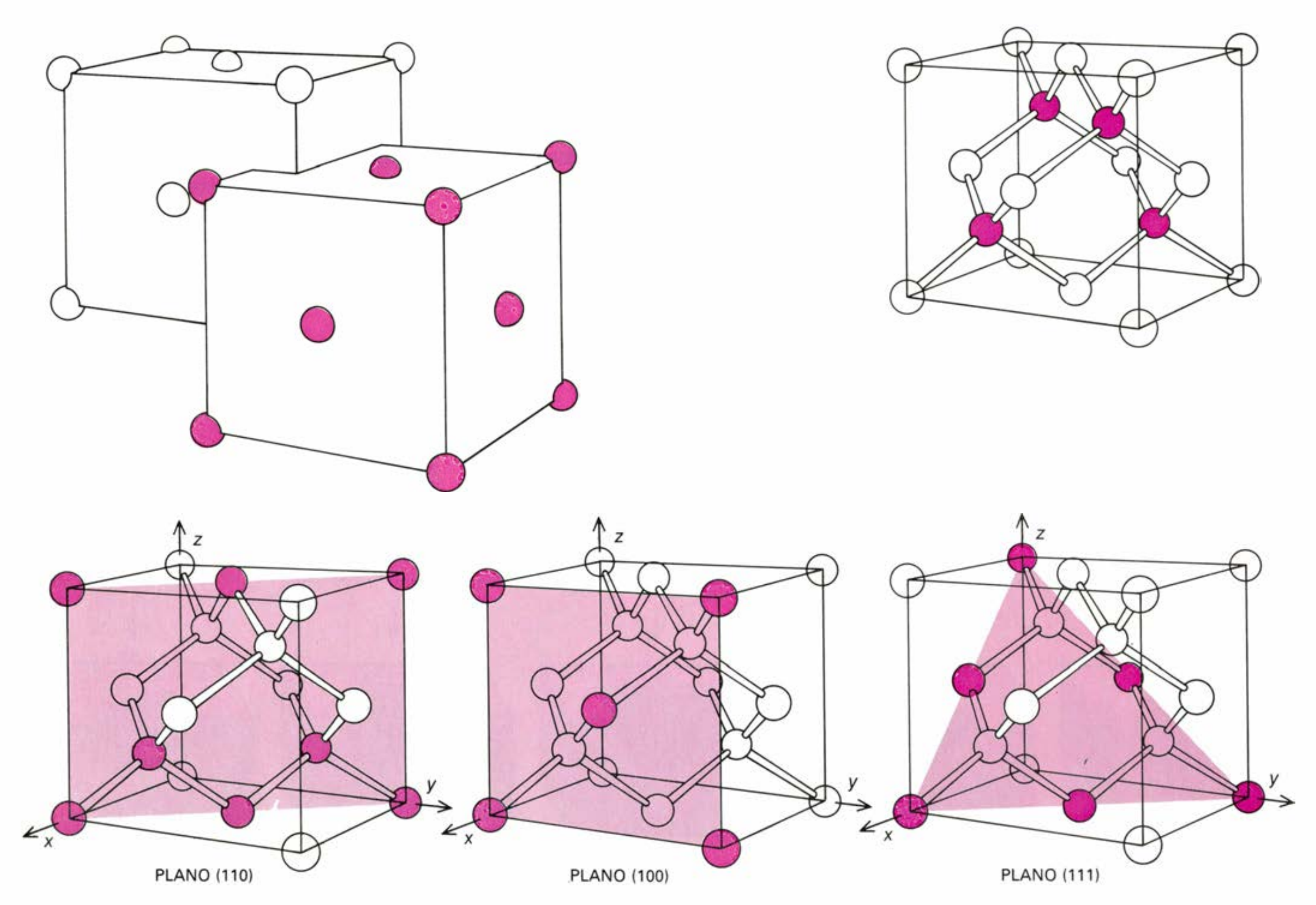
En la micromecanización del silicio se utilizan dos tipos de grabadores. Los llamados grabadores isotrópicos atacan al cristal de silicio con la misma intensidad en todas direcciones y crean formas redondeadas. Los grabadores no isotrópicos, conocidos también como grabadores cristalográficos dependientes de la orientación, atacan al cristal con intensidades variables según las distintas direcciones de la red cristalina; las figuras que originan estos grabadores presentan formas muy bien definidas, con bordes delimitados. Los grabadores isotrópicos más utilizados están constituidos por mezclas de ácidos

fluorhídrico, nítrico y acético (conocidos como grabadores HNA). Los grabadores anisotrópicos son todos ellos soluciones alcalinas calientes: la disolución acuosa del hidróxido potásico (KOH) o hidróxido sódico (NaOH) y una mezcla de etilenodiamina, pirocatecol y agua conocida por EDP.

En la micromecanización se utilizan más los grabadores anisotrópicos que los isotrópicos. Una operación de micromecanización típica, el grabado de una fila de agujeros que atraviesen totalmente el disco base, nos proporciona la explicación de este hecho. Un grabador isotrópico se mueve a partir de una abertura en el recubrimiento del óxido, ensanchando el pozo grabado al mismo tiempo que profundiza en él. El pozo en cuestión es al menos dos veces más ancho que profundo; las aberturas en el óxido estarán separadas entre sí una distancia (medida de borde a borde) que sea al menos igual al grosor del disco, si se pretende perforar con varios agujeros la oblea de silicio. Por su parte, si se utilizan correctamente, los grabadores anisotrópicos originan pozos con paredes bien definidas, cuya profundidad se puede aumentar sin que aumente su anchura. Por consiguiente, un grabador anisotrópico permite obtener filas de agujeros estrechamente espaciados entre sí; los bordes de las aberturas en la capa de óxido pueden estar tan próximos como permita la fotolitografía.

La forma del agujero que origina un grabador anisotrópico depende de la orientación de los planos atómicos en la oblea de silicio. De la misma forma que una serie de filas monodimensionales de tallos de maíz se puede identificar bajo ángulos diferentes como un campo bidimensional, varios planos de átomos bidimensionales se sitúan según ángulos diferentes en un cristal tridimensional. Los planos de átomos en el cristal se pueden identificar por medio de un sistema de coordenadas asignado al cristal.

Cada átomo de silicio se enlaza con cuatro átomos próximos. La estructura de los enlaces es tetraédrica, es decir, cada átomo está situado en el centro de un tetraedro definido por los cuatro átomos a los que dicho átomo está unido. Esta estructura resulta inadecuada como base de un sistema coordinado, ya que los enlaces atómicos no forman ángulos rectos entre sí. Es más conveniente considerar que un cristal de silicio está constituido por una serie de capas superpuestas de cubos idénticos. Cada cubo posee un átomo en cada vértice y en el centro de cada cara y está entrelazado con los cuatro cubos más próximos. Esta estructura cúbica centrada en las caras puede servir como base de un sistema de coordenadas. Cualquier átomo se puede designar co-



3. CRISTALES DE SILICIO, que poseen la misma estructura cristalina que el diamante, denominada cúbica centrada en las caras entrelazadas. En cada uno de los vértices y en el centro de cada una de las caras del cubo unitario se sitúa un átomo; los cubos se conectan entre sí de tal forma que en cada unidad se encuentran varios átomos de los cubos vecinos. Los ejes del cubo unitario definen un sistema de coordenadas rectangulares que permite especificar las direcciones y los planos existentes dentro del cristal. Una dirección cristalina viene determinada por tres coordenadas conocidas como índices de Miller, cuyos valores son múltiplos enteros de la longitud de una arista del cubo unitario. El mismo conjunto de índices sirve para designar los planos perpendiculares a dicha dirección. La orientación de los cristales constituye un dato de la máxima importancia en la fabricación de dispositivos micromecánicos, habida cuenta de que algunos de los grabadores utilizados actúan con intensidades diferentes según las distintas direcciones del cristal. La mayoría de tales grabadores anisotrópicos progresan rápidamente a lo largo de la dirección del cristal perpendicular al plano (110) y menos rápidamente en la dirección perpendicular al plano (100). La dirección perpendicular al plano (111), abajo a la derecha, se ve atacada muy poco, si es que lo es en alguna cuantía.

mo el origen de un sistema de coordenadas cuyos ejes coinciden con las aristas del cubo al que pertenece dicho átomo. La unidad básica de medida es la longitud de la arista del cubo. En el cristal, un vector o una dirección se describe mediante un conjunto de tres coordenadas conocidas por índices de Miller. Así, el vector [110] cruza diagonalmente una de las caras del cubo unitario. Las coordenadas de un vector dado se pueden designar también mediante un conjunto de planos atómicos perpendiculares a dicha dirección. Concretamente, la notación (100) describe todos los planos perpendiculares al eje *x*. Merced a la simetría de la estructura del cristal, las direcciones *x*, *y* y *z* son intercambiables y se puede introducir una notación generalizada para describir todas las direcciones equivalentes y conjuntos de planos. La notación <110> designa las diagonales a lo largo de cualquier cara del cubo unidad; la notación {110} designa el conjunto de planos perpendiculares a todos los vectores <110>. Los diferentes corchetes y paréntesis sirven para diferenciar entre los planos y las direcciones y entre los planos y direcciones generalizados y los planos y direcciones particulares.

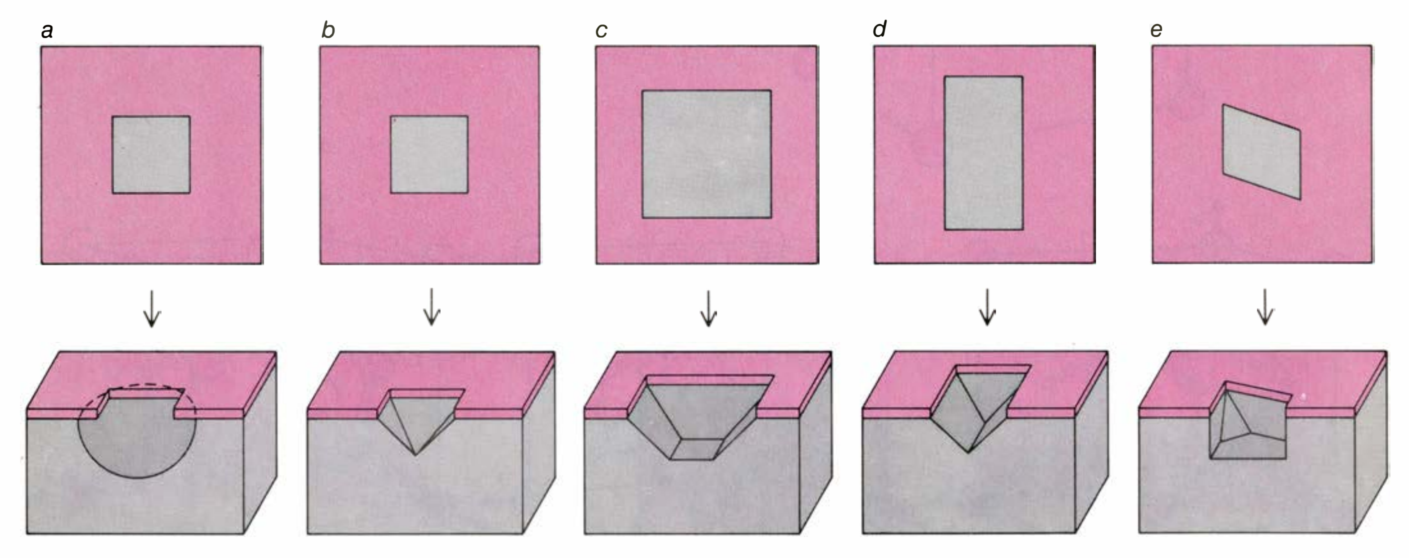
Desconocemos los mecanismos que condicionan la dependencia de la velocidad de grabado de un grabador no isotrópico con la dirección del cristal. Las diferencias en la velocidad de grabado dependen de la temperatura y de la composición del grabador y parecen estar relacionadas también con la densidad de enlaces atómicos existentes en el plano del silicio expuesto a su acción y con el radio de los iones hidroxilos hidratados en la solución de grabado.

Sin embargo, la experiencia ha demostrado que los planos {111} se graban muy lentamente, si es que lo son en algún grado, por parte de los grabadores anisotrópicos. Por el contrario, los planos {110} se suelen grabar muy rápidamente; los planos {100} lo hacen con velocidad intermedia. Los agujeros sencillos sólo pueden tener facetas {111} y {100}. Otros planos de orden superior, así los planos {221}, se graban también con velocidad media y pueden aparecer como meras facetas de formas más complejas.

El tipo de agujero que un grabador anisotrópico forma en una oblea de silicio viene determinado por la orientación cristalina en su superficie, así como por la dependencia del grabador con respecto a la orientación del cristal. Los dos tipos de obleas de silicio más utilizadas en el grabado no isotrópico son las láminas <100> (donde los planos {100} son paralelos a la superficie y una dirección <100> es perpendicular a ella) y las obleas <110> (donde los planos {110} son paralelos a la superficie y una dirección <110> es perpendicular a ella).

La orientación, el tamaño y la forma de la abertura del óxido superficial de una oblea de silicio interviene también de un modo destacado en la determinación del tipo de agujero formado, tal como ilustra el siguiente ejemplo. Si se practica un agujero cuadrado en una capa de óxido sobre una oblea de tipo <100>, el tratamiento con un grabador anisotrópico puede crear un pozo en forma de pirámide invertida. Cuatro conjuntos de planos {111} interseccionan con la superficie de una oblea <100> a lo largo de las dos direcciones perpendiculares <110> que existen en el plano superficial. Durante el grabado anisotrópico, el plano de la superficie expuesta {100} se graba en profundidad a una velocidad constante; el pozo que se produce al principio del proceso de grabado tiene el fondo plano. Las paredes del pozo están constituidas por cuatro planos {111} (formando ángulos de 55 grados con la superficie). A medida que el proceso de grabado avanza, nuevos planos {111} resultan expuestos y la zona del fondo plano se encoge. Con el tiempo, los planos {111} se cortan en un punto; cuando tal ocurre, ha desaparecido el fondo plano al que nos estamos refiriendo. Al llegar a esta situación se interrumpe el grabado, pues no restan planos susceptibles de ser atacados.

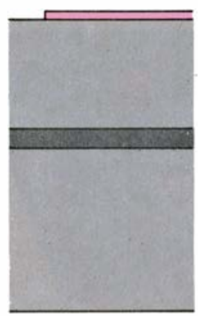
La abertura cuadrada del óxido debe estar alineada con una precisión de uno



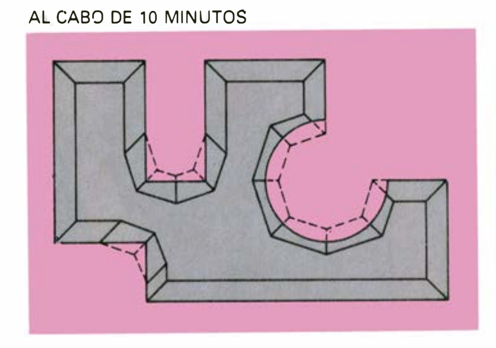
4. LOS GRABADORES ANISOTROPICOS originan agujeros cuyas formas concretas vienen determinadas por la posición de los planos cristalinos que son poco atacados por dichas sustancias, al contrario de lo que sucede con los grabadores isotrópicos que siempre producen agujeros redondeados (*a, arriba, el primero empezando por la izquierda*). De hecho, la forma de un agujero atacado anisotrópicamente no sólo depende de la orientación del cristal en la superficie de la oblea de silicio, sino también de la forma y orientación de las aberturas practicadas en el recubrimiento de dicha superficie y de la relación del propio grabador con esa misma orientación. Una abertura cuadrada orientada a lo largo de las direcciones <110> de una oblea <100> da lugar a un pozo piramidal con paredes {111} (*b*). Si la abertura es mayor, el punto de intersección de los planos {111} será más profundo y, si se detiene el ataque antes de que se alcance la superficie trasera de la oblea, se obtendrá un pozo con fondo plano (*c*). Una abertura rectangular sobre esta misma oblea proporciona un canal en forma de *V* (*d*). Atacando una oblea con una orientación superficial diferente, se crean una serie de agujeros con paredes paralelas. En una oblea <110>, dos conjuntos de planos {111} son perpendiculares a la superficie, aunque no lo sean entre sí. Cuando se orienta adecuadamente una abertura de óxido sobre una oblea <110>, la grabación da lugar a un agujero con paredes verticales. Las paredes que se cortan según ángulos agudos se unen entre sí por medio de otros planos. Todas estas formas representan algunas de las muchas variantes que consigue la técnica de grabado químico.

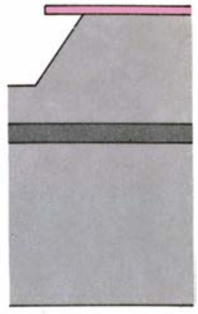
ANTES DEL GRABADO



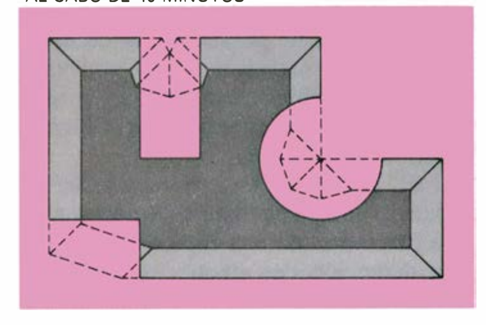


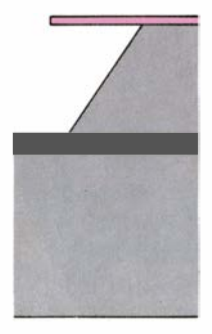
AL CABO DE 10 MINUTOS



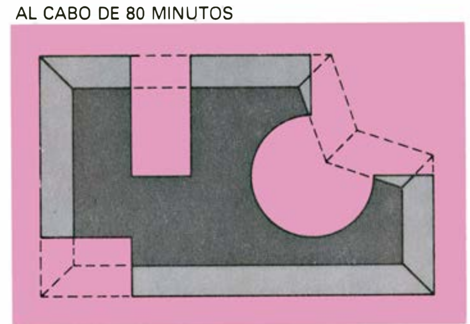


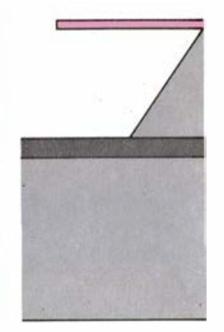
AL CABO DE 40 MINUTOS





AL CABO DE 80 MINUTOS





5. PUENTES VOLADIZOS DE OXIDO construidos socavando una capa de óxido mediante un ataque químico. Las formas convexas de óxido se pueden socavar con la ayuda de un grabador anisotrópico; sin embargo, esto no es posible con las formas cóncavas orientadas adecuadamente. La abertura de óxido rectangular representada en esta figura posee una forma cóncava; los apéndices de óxido que sobresalen en la abertura tienen formas convexas. Si el avance en profundidad del grabador se limita con la ayuda de una capa de detención del ataque, un grabador anisotrópico prolongado permitirá obtener un apéndice de óxido colgando sobre un pozo rectangular poco profundo con paredes {111} y fondo plano, que coincidirá con la capa de detención del ataque. Los puentes voladizos que se forman al socavar totalmente la capa de óxido del extremo superior izquierdo se han utilizado para fabricar acelerómetros, conjuntos de interruptores controlados electrostáticamente y baterías de espejos. La excavación dificulta la grabación de un agujero en el que se proyecta una esquina de silicio en ángulo recto. El silicio situado bajo esta esquina resulta redondeado en el transcurso del proceso de grabado. Para producir agujeros donde los bordes formen ángulos rectos, puede recurrirse también al ataque químico con tiempos controlados de figuras circulares.

o dos grados respecto a las direcciones <110> de la superficie de la oblea, con el fin de obtener pozos piramidales que se correspondan exactamente con el recubrimiento de óxido en lugar de quedar girados respecto a él. La mayoría de las obleas de silicio tienen un borde plano paralelo a una de las direcciones <110>. El recubrimiento cuadrado está alineado con este borde plano durante el proceso de fotolitografía que precede a la operación de grabado.

El tamaño de la figura dibujada en el óxido determina, además de la superficie del pozo, su profundidad. En efecto, cuanto mayor es la abertura del cuadrado de óxido más profundo se encuentra el punto en el que los planos {111} interseccionan entre sí. Si la abertura del óxido es bastante grande, los planos {111} no se cortan en el seno de la oblea. En consecuencia, el pozo grabado alcanza la cara trasera de la oblea, creando pequeñas aberturas cuadradas en la superficie posterior.

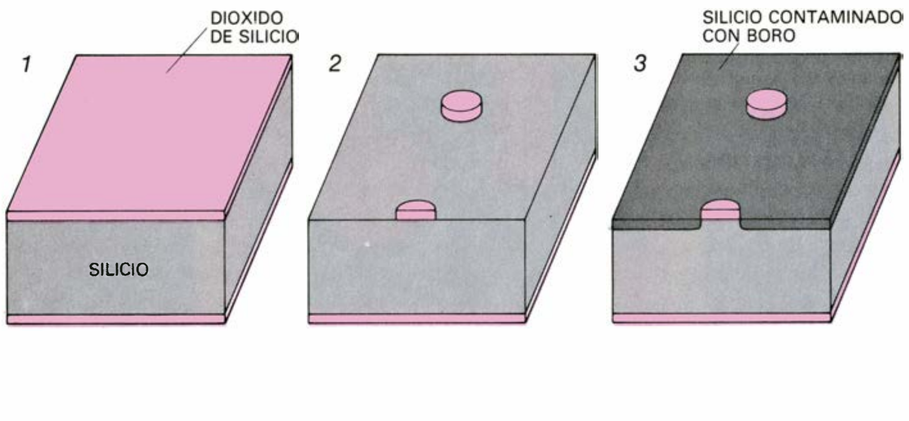
La influencia de la forma de la abertura del óxido sobre la geometría del pozo grabado es evidente. Si la abertura en el óxido depositado sobre una oblea <100> es rectangular y no cuadrada, el grabado crea una zanja larga y no una pirámide invertida. Las paredes y los extremos de la zanja son planos {111}; si el grabado se completa, la sección transversal de la zanja adquiere forma de *V*.

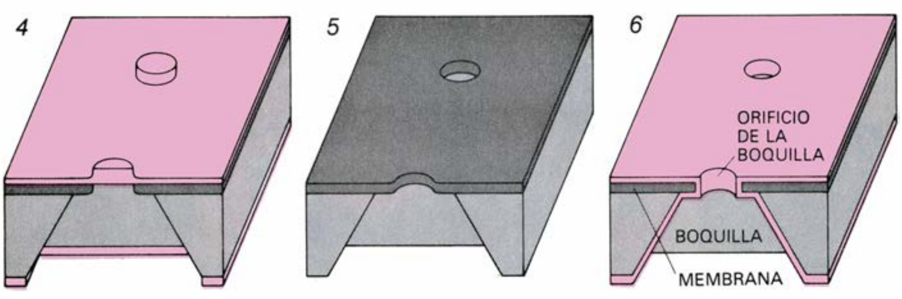
Los pozos piramidales, las zanjas en forma de *V* y otros tipos de pozos y agujeros constituyen otros tantos elementos estructurales de los que se componen los dispositivos micromecánicos. Una segunda unidad estructural es una membrana delgada de silicio. Una membrana se puede formar atacando una oblea de silicio durante un período de tiempo ligeramente menor que el necesario para atravesarla. Sin embargo, utilizando este método resulta muy difícil obtener diafragmas de espesor uniforme, dado que el espesor del diafragma viene determinado por el de la oblea y los espesores de ésta varían al menos en 10 micrometros a lo largo de una misma y de una oblea a otra.

Las membranas con espesores y formas más precisos se obtienen mediante una técnica que se basa en otra propiedad de los grabadores anisotrópicos. La velocidad con que las soluciones de grabado atacan a una oblea de silicio no sólo depende de la orientación del cristal, sino también de la cuantía con que el silicio se ha cargado con átomos de impurezas tales como boro, fósforo, arsénico o antimonio.

La velocidad de grabado de un grabador dependiente de las impurezas (una categoría intermedia entre los grabadores isotrópicos y los anisotrópicos) está limitada por el tipo de impurezas y de su concentración. El grabador isotrópico HNA depende de las impurezas en ciertas mezclas; esta sustancia graba el silicio fuertemente cargado con impurezas mucho más deprisa que el silicio con pocas impurezas. Por otra parte, los grabadores anisotrópicos, así el hidróxido de potasio y el EDP, atacan el silicio fuertemente cargado con boro mucho más lentamente de lo que lo hacen con el silicio poco cargado con boro.

El primer paso para crear una membrana por grabado anisotrópico consiste en desarrollar una zona de detención del proceso de grabado, constituida por una capa de silicio fuertemente cargado con boro, situada en la superficie de la oblea. El proceso se inicia depositando





6. BOQUILLA DE TINTA construida en una oblea de silicio y destinada a una impresora de alta velocidad. La boquilla tiene la forma de un pozo piramidal cuyo fondo está constituido por una membrana delgada a través de la cual se ha perforado un orificio circular. El diámetro de este orificio viene determinado por las tres primeras etapas del proceso de grabado. La oblea se recubre de óxido y se graban ciertos puntos de una sustancia protectora fotosensible sobre la superficie. A continuación, se elimina el óxido que está protegido por la sustancia fotosensible y se obtiene una imagen de puntos de óxido. Cuando el boro se difunde en el silicio, se crea una capa de silicio contaminado por boro a lo largo de toda la oblea excepto bajo dichos puntos. Se practica una abertura cuadrada en el óxido de la parte inferior de la oblea y se somete al ataque de un grabador anisotrópico. El grabador crea un pozo piramidal con paredes {111}; el ataque se detiene al llegar a la capa contaminada. Sin embargo, el ataque continua a través de los puntos no contaminados, de forma tal que, cuando se elimina todo el óxido de los orificios de la boquilla, atraviesan la oblea. Se vuelve a oxidar con el fin de que todas las superficies queden recubiertas por una capa de protección. Se fabrican varias filas de boquillas sobre una misma lámina.

una capa de alguna sustancia rica en boro sobre la superficie delantera de la lámina; a continuación, la oblea se calienta entre 1000 y 1200 grados Celsius. A estas temperaturas, los átomos de boro se difunden en el silicio. La profundidad de la capa de carga se controla variando la temperatura y la duración del proceso de difusión. Si la lámina de silicio se trata, por su parte trasera, con un grabador anisotrópico, el grabador la ataca hasta llegar a la capa delgada, y allí se detiene, creando una membrana cuyo espesor coincide con el de la capa de las impurezas presentes en la superficie de la lámina.

Si el diseñador de estos dispositivos desea que la membrana en cuestión esté situada un tanto por debajo de la superficie de la oblea, procede a cargar su superficie de forma similar a como se ha explicado anteriormente; a continuación, desarrolla una capa de silicio sobre la capa de impurezas calentándola en una atmósfera de gas silano (SiH_4). Este gas se descompone a altas temperaturas; se produce un depósito de silicio sobre la superficie de la lámina. Esta nueva capa, llamada capa epitaxial, adquiere la misma estructura y orientación que el cristal de silicio sobre el que está depositada. El espesor de la capa epitaxial suele ser del orden de 5 a 20 micrometros.

Si la oblea se ataca por ambas caras con un grabador anisotrópico, se crea una membrana a una profundidad igual al espesor de la capa epitaxial. El espesor de la membrana viene determinado, como siempre, por el espesor de la capa de impurezas de boro.

Un tercer ejemplo de estructura micromecánica es el puente voladizo, constituido por una barra delgada de dióxido de silicio apoyada solamente por uno de sus extremos. En muchas aplicaciones este puente está suspendido sobre un pozo practicado en la oblea de silicio. Las técnicas para construir el puente y el pozo fueron desarrolladas por un grupo de trabajo dirigido por Kurt E. Petersen, del Laboratorio de Investigación San José de la International Business Machines Corporation (IBM). Estas técnicas se basan en dos propiedades de los grabadores anisotrópicos: su dependencia con respecto a las impurezas y su tendencia a socavar las capas de óxido según formas convexas.

Las aberturas del óxido a las que nos hemos referido anteriormente tienen formas cóncavas, es decir, la línea que separa el óxido de la abertura es siempre cóncava con respecto a la abertura. Una forma de óxido cóncava no la socava ningún grabador anisotrópico si la abertura está orientada correctamente sobre la superficie de la oblea. Ahora bien, si a partir de uno de los lados de tal abertura se proyecta una cinta de óxido, sus extremos serán convexos respecto al silicio expuesto. Un grabador anisotrópico sí socava estas formas convexas del óxido.

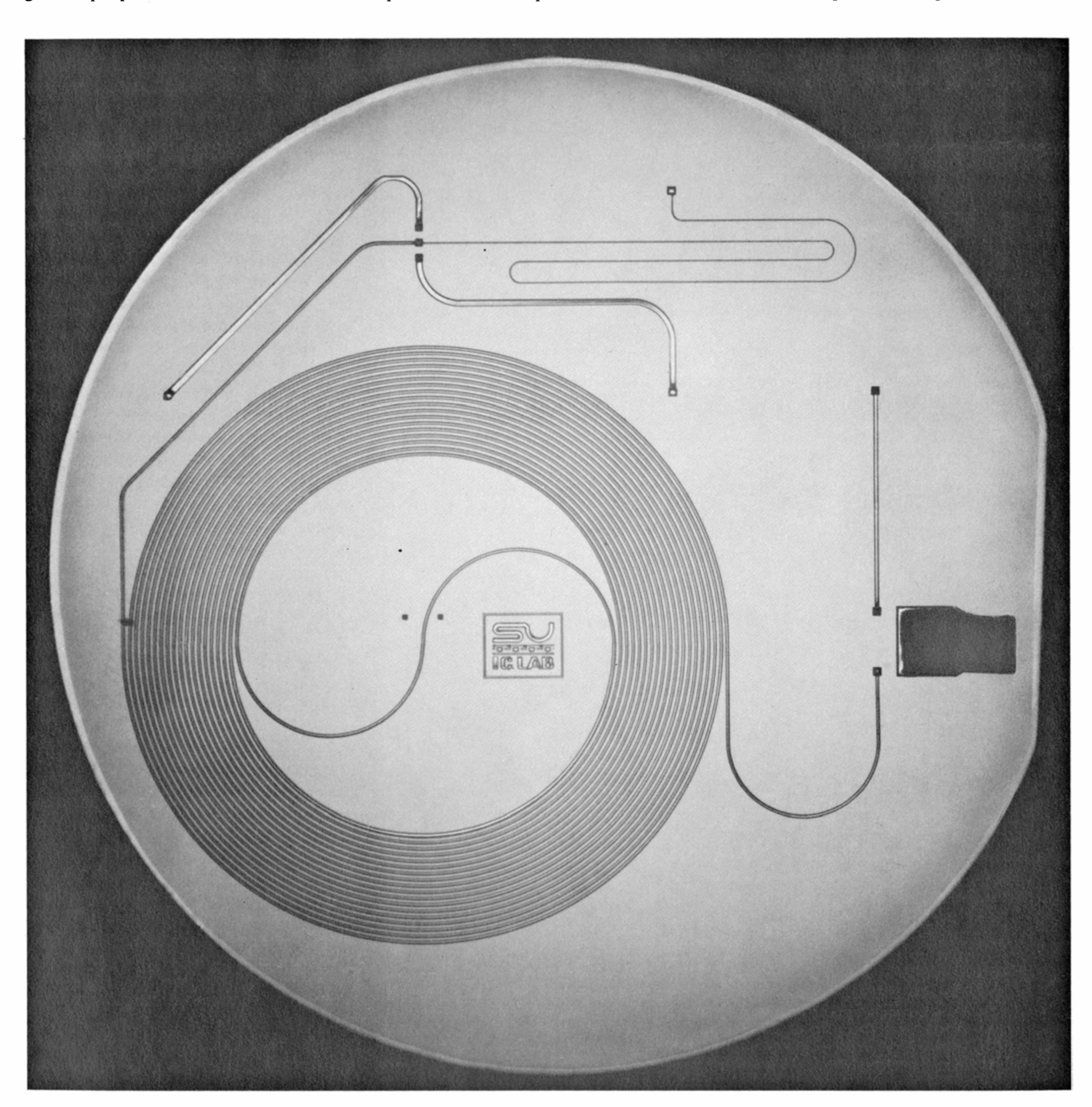
El primer paso para construir un puente sobre un pozo hueco consiste en la formación de una capa de boro para la detención del ataque. A continuación, se deposita una capa de silicio epitaxial sobre la capa con impurezas de boro, se produce una capa de óxido sobre el silicio epitaxial y se practica una abertura en el óxido. La lámina se somete entonces al ataque de un grabador anisotrópico. Si en la abertura del óxido se coloca una cinta de óxido, el grabador empieza a socavarlo. En el extremo fijo de la cinta de óxido el grabador acaba por encontrar un plano {111} y se detiene el ataque. Dado que la capa de detención limita el proceso del ataque hacia abajo, el resultado final de la operación es un puente de óxido que cruza sobre una depresión hueca en el sustrato.

El trabajo en el campo de la micromecanización se inició en el año 1965 en el Laboratorio de Circuitos Integrados de la Universidad de Stanford, cuando Kensall D. Wise (hoy en la Universidad de Michigan) y uno de los autores de este artículo (Angell) construyeron una fila de electrodos, apenas distantes entre sí, con el fin de medir potenciales eléctricos en el cerebro. El proyecto permitió obtener una serie muy amplia de sondas cerebrales. Desde aquel entonces, los trabajos de micromecanización de Stanford han desarrollado sensores de presión, acelerómetros, cromatógrafos de gases, termómetros en miniatura y estructuras para el aislamiento dieléctrico de circuitos microelectrónicos. En la actualidad, las técnicas y dispositivos micromecánicos son objeto de estudio y aplicación en otras universidades y laboratorios industriales. Los dispositivos que

vamos a describir a continuación incluyen algunos de los desarrollados por nosotros y por otros investigadores.

Las filas de boquillas fabricadas con silicio han encontrado aplicación en las impresoras por chorro de tinta utilizadas en los ordenadores digitales. En una impresora de este tipo, se obliga a la tinta a salir por una fila de boquillas, lo que produce una serie de chorros paralelos que se convierten en gotitas a medida que salen de las boquillas. Las gotitas que proceden de cada uno de los chorros pasan junto a unos electrodos y se cargan eléctricamente o permanecen eléctricamente neutras. Una gotita sin carga no sufre deflexión alguna cuando atraviesa un campo eléctrico y produce un punto sobre un papel, pero si la gotita está cargada es desviada por el campo y, en lugar de llegar al papel, queda interceptada por un canal de recogida. La impresión de un carácter determinado se obtiene seleccionando las gotitas procedentes de una fila vertical de boquillas a medida que todo el mecanismo se mueve horizontalmente a lo largo del papel.

Ernest Bassous, Larry Kuhn y sus colegas, del Centro de Investigación Thomas J. Watson de IBM, llevaron a cabo los primeros trabajos sobre estos dispositivos micromecánicos, desarrollando muchos sistemas diferentes de boquillas. La forma más sencilla de construir estas boquillas consiste en practicar un pozo piramidal en el silicio. Si el punto de intersección de los planos {111} que forman las paredes del pozo está situa-



7. EN EL CROMATOGRAFO DE GASES DE SILICIO tenemos un instrumento construido por entero en una oblea de cinco centímetros de diámetro. La columna capilar del cromatógrafo está constituida por una espiral de 1,5 metros de longitud. Esta columna se obtiene grabando un surco en la oblea de silicio y cubriéndolo posteriormente con un vidrio. El prototipo expuesto en la figura se desarrolló por especialistas de la Universidad de Stanford.

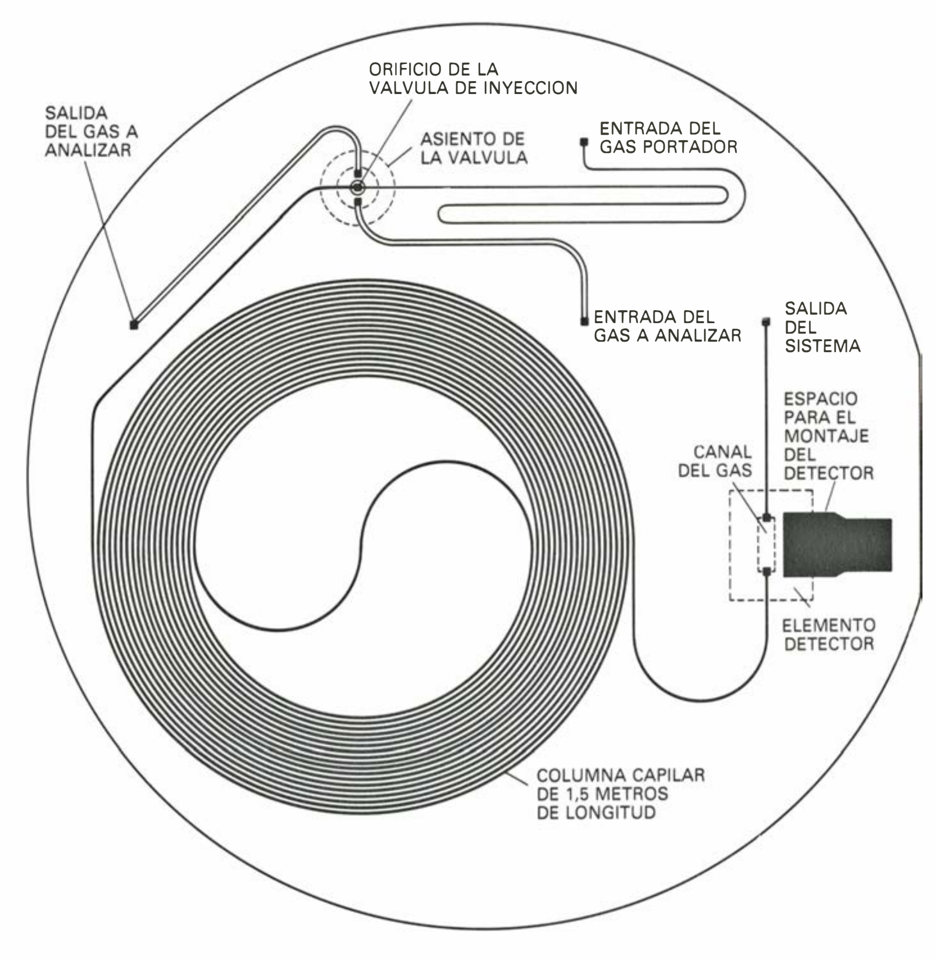
do por debajo de la superficie del fondo de la lámina se forma una pequeña abertura cuadrada. Sin embargo, con esta técnica es difícil producir orificios de tamaño uniforme, en parte debido a que las dimensiones del orificio dependen del espesor de la lámina; tal como señalamos antes, no es fácil controlar ese espesor.

¿Cómo resolver el problema? Colocando una membrana de tres micrometros de espesor en el fondo de cada cavidad de boquilla. En esta lámina se graba un pozo piramidal, que sirve de cavidad principal, pero el orificio de la boquilla viene definido por un procedimiento de difusión de boro y no por la geometría del pozo. Utilizando este método, Bassous y sus colaboradores han construido unas filas con ocho boquillas de orificios circulares de 20 micrometros de diámetro y espaciadas entre sí 0,3 milímetros exactos.

El primer paso en la construcción de estas filas consiste en determinar los orificios de las boquillas como puntos de una sustancia protectora fotosensible sobre la superficie delantera de una oblea de silicio recubierta de óxido por ambos lados. Todo el óxido de la superficie delantera es eliminado químicamente, con la excepción de los puntos protegidos por la sustancia fotosensible. A continuación, la superficie de las láminas u obleas se impurifica con boro, hasta una profundidad de tres micrometros, calentando dichas láminas en un horno de difusión. Unicamente el silicio situado bajo los puntos de óxido permanece sin contaminar por el boro.

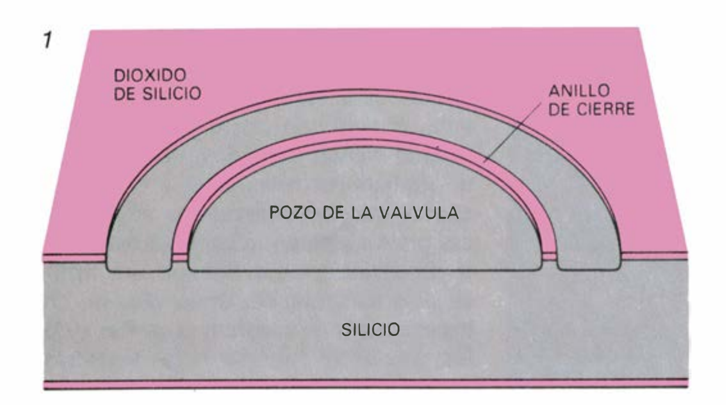
Más tarde, mediante un proceso de fotolitografía y grabado, se define una figura de aberturas cuadradas en el óxido que cubre la superficie trasera de la oblea. Las aberturas se centran bajo los puntos no contaminados de la superficie delantera. Se somete luego la oblea al ataque de un grabador anisotrópico dependiente de la contaminación, que elimina hasta el 99 por ciento del espesor de aquella, pero se detiene al llegar a la capa contaminada con boro. Sin embargo, el ataque prosigue normalmente a través de los puntos no contaminados, con lo cual se crean los orificios de las boquillas. El diámetro de cada orificio, el espesor de la membrana que los rodea y el emplazamiento de unos orificios respecto a otros se controlan con una precisión del orden de micrometros: todas estas características no están definidas por el espesor de la oblea o la duración del proceso de grabado, sino por la fotolitografía y la difusión del boro.

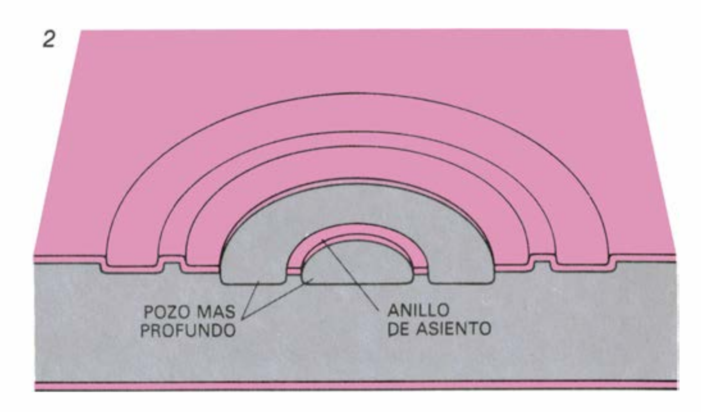
El dispositivo micromecánico más refinado que se ha construido hasta ahora es un cromatógrafo de gases del tamaño de una caja de cerillas. Dos de los autores (Terry y Angell) fabricaron un prototipo de dicho cromatógrafo en 1975. La empresa Microsensor Technology, Inc. la desarrolla hoy a escala comercial.

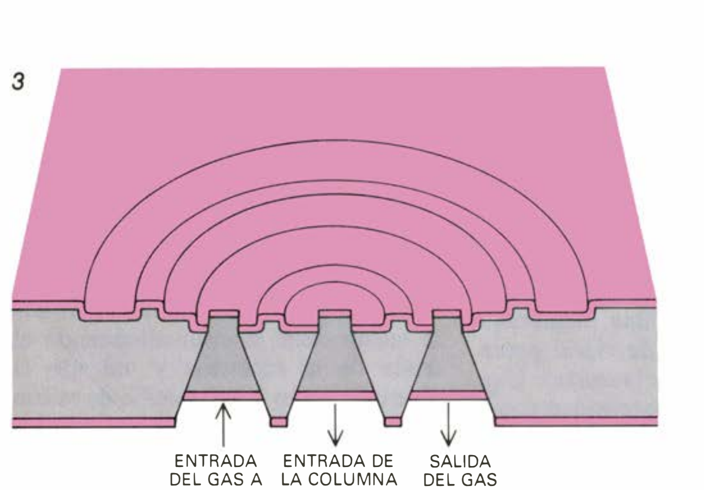


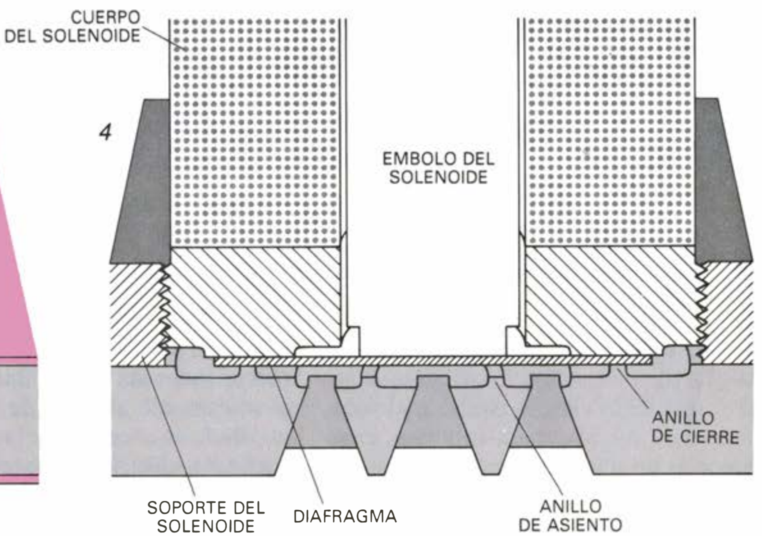
8. SEPARACION DE LOS GASES en el cromatógrafo de silicio. Dicho proceso se basa en las diferencias de solubilidad de los gases que hay en el líquido que recubre las paredes de una columna capilar. Un gas de transporte inerte fluye continuamente a través de dicho conducto capilar. Cuando se abre una válvula, penetra en la columna un impulso del gas a analizar arrastrado por el gas portador. A medida que los gases de la muestra avanzan a lo largo de la columna se van adsorbiendo y desadsorbiendo, por la delgada capa del líquido que recubre las paredes de la columna. Cada gas se identifica por su tiempo de retención en la columna. A medida que los gases que constituyen cierta mezcla llegan al final de la columna, van atravesando de un agujero a un canal situado en la parte posterior de la oblea de silicio, donde se ha montado un detector de conductividad térmica. La conductividad térmica de los gases de la muestra es menor que la del gas portador (helio) y se producen máximos de voltaje en la señal de salida del detector. El volumen de cada gas viene determinado por el área encerrada en el correspondiente pico de voltaje.

La cromatografía de gases es una técnica analítica para separar, identificar y medir la cantidad de los gases que constituyen una mezcla. Se inyecta una muestra de dicha mezcla a través de una válvula en una larga columna capilar, a lo largo de la cual es arrastrada por un gas de transporte inerte que, en la mayoría de los casos, suele ser helio. Las paredes de la columna capilar están recubiertas por una capa delgada de un material, aceite de silicona o un polímero, en el que los diferentes gases que constituyen la mezcla tienen grados de solubilidad distintos. A medida que los componentes de la mezcla atraviesan la columna son adsorbidos y desadsorbidos por la sustancia de recubrimiento. Dado que el tiempo en que cada uno de los componentes de la mezcla permanece adsorbido depende de su solubilidad, la velocidad con que avanza cada gas en el seno de la columna es diferente y, como consecuencia, cada gas emerge de la columna en un instante diferente. La corriente de salida pasa por un detector que mide alguna propiedad del gas, verbigracia su conductividad térmica. La señal de salida que se genera en el detector consiste en una serie de picos en forma de campana separados por regiones planas que corresponden al gas inerte de transporte. Muchos cromatógrafos de gases modernos poseen un microordenador aso-









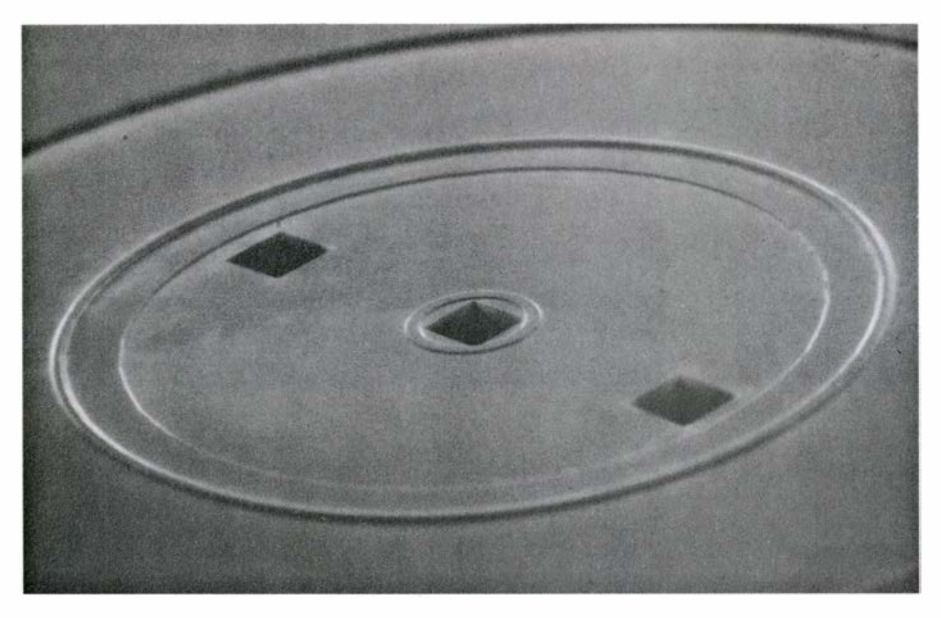
9. VALVULA DE INYECCION de la muestra del cromatógrafo de gases en miniatura, construida sobre la misma oblea de silicio que la columna capilar. Dos etapas de grabado isotrópico definen el anillo de cierre y el asiento de la válvula. Un proceso de grabado anisotrópico define posteriormente los orificios de alimentación en el pozo de la válvula. En la primera fase del grabado isotrópico, se crea un anillo de cierre de bordes muy afilados que sobresale unos 10 micrometros sobre el fondo del pozo. En la segunda fase se obtiene un pozo concéntrico menor que el anterior, provisto de otro anillo –el asiento de la válvula– excavado 10 micrometros en el fondo del pozo. En la tercera fase se crean los orificios de alimentación mediante un grabado anisotrópico a través de ciertas aberturas cuadradas practicadas en la capa de óxido que recubre la superficie posterior de la lámina de silicio. El agujero situado dentro de la válvula comunica con la columna capilar; los otros dos agujeros comunican con los canales del gas a analizar. Un diafragma de metal y teflón y un solenoide completan la válvula (*esquema número 4, abajo a la derecha*). El diagrama comprime el asiento de la válvula, impidiendo en lo posible que el gas a analizar llegue a la columna. Cuando interviene el solenoide, el diafragma se relaja y el gas fluye entonces a través del asiento de la válvula.

ciado, donde se identifica cada gas ajustando su tiempo de retención con los tiempos de retención conocidos y se mide la cantidad que existe de cada gas calculando la superficie encerrada bajo el pico correspondiente.

Al igual que un cromatógrafo de gases clásico, el modelo miniaturizado consta de una columna capilar, una válvula para la inyección de las muestras y un detector de conductividad térmica. Sin embargo, todas las partes del cromatógrafo de gases en miniatura están incorporadas en una sola oblea de silicio de cinco centímetros de diámetro. El volumen de la columna capilar es mucho menor que el de una columna capilar convencional. Debido a que un cromatógrafo de gases opera adecuadamente en el caso exclusivo de que el volumen de la muestra de gas inyectada sea mucho menor que el volumen de la columna, hubo que diseñar una válvula de inyección en miniatura para acompañar la columna capilar en miniatura. Con el fin de que el volumen interno entre la válvula y la columna y entre la columna y el detector sea el menor posible, la válvula y la columna se construyen sobre la misma lámina y el detector está constituido por un pequeño fragmento de silicio montado sobre la oblea en cuestión.

La columna capilar está formada por un canal de 1,5 metros de longitud, cuya configuración espiral posibilita su inclusión en la lámina. Una placa de vidrio adherida a la lámina forma la parte superior de la columna. Un agujero situado en el extremo de entrada de la columna conduce a la superficie inferior de la lámina u oblea. El gas helio de transporte penetra en la columna a través de este agujero. Un poco más allá existe otro agujero que conduce a una válvula situada sobre la superficie trasera de la oblea. La muestra de gas a analizar penetra por un canal independiente a través de otro orificio, fluye por ese canal hasta llegar a la válvula y es inyectado por dicha válvula en la columna capilar.

El asiento de la válvula descansa sobre la superficie trasera de la oblea; consta de un anillo de cierre de silicio, que rodea tanto el orificio de entrada como el de salida, y un anillo de asiento de silicio, que rodea solamente al orificio de salida que conduce a la columna capilar. Un diafragma de níquel y teflón está encajado en el anillo de cierre. Junto a dicho diafragma suele colocarse un solenoide sujetado con un muelle. Cuando el solenoide entra en acción, el diafragma se abre y permite que el gas



10. MICROFOTOGRAFIA ELECTRONICA del fondo de la válvula. Tiene cuatro milímetros de diámetro y está limitado por dos anillos concéntricos de silicio: el cierre de la válvula (anillo exterior) y el asiento de la válvula (anillo interior). Las paredes del anillo de 50 micrometros de espesor. El orificio practicado en el centro conduce a la columna capilar; los otros dos orificios, a los canales de admisión y de salida.

circule desde el orificio de entrada hasta el orificio de salida. El volumen inactivo real de la válvula equivale al volumen del orificio capilar-columna, cuyo valor es de unos cuatro nanolitros.

En el extremo de salida de la columna capilar existe otro orificio que comunica con un canal de conducción grabado en la superficie trasera de la lámina de silicio. La unidad sobre la que se ha construido el detector de conductividad térmica está situada en este canal y se encuentra incorporada en la oblea. El detector consiste en una resistencia montada sobre una membrana térmicamente aislante de vidrio pyrex instalada en el centro de la unidad. Una corriente eléctrica de intensidad constante atraviesa el citado resistor. La conductividad térmica de los gases a analizar es menor que la del gas de transporte y origina máximos en el correspondiente voltaje. La amplitud de los respectivos picos es proporcional a la cantidad de cada uno de los gases presentes en la mezcla.

La micromecanización de la oblea de silicio que constituye la base del cromatógrafo se inicia con la grabación del asiento de la válvula. El pozo de la válvula, el anillo de cierre y el anillo del asiento vienen definidos por el ataque de grabadores isotrópicos a través de ciertas aberturas circulares concéntricas practicadas en la capa de óxido de la cara trasera de la lámina. Los orificios de alimentación de la válvula, de los extremos de la columna capilar y de los canales de entrada de las muestras de gas a analizar se forman por grabación anisotrópica a través de aberturas de forma cuadrada definidas en la capa de óxido que recubre la superficie delantera de la lámina. Finalmente, la columna capilar y los canales de gas de transporte se obtienen por ataque isotrópico a través de las figuras de óxido grabadas en la superficie delantera de la lámina.

La columna capilar está constituida por un canal en forma de espiral situado a unos 40 micrometros de profundidad bajo la superficie de la lámina base de silicio. La espiral abierta se convierte en un canal interior eliminando el óxido de la superficie y uniendo la oblea de silicio a una lámina de vidrio pyrex. Esta técnica de fijación, conocida por unión anódica, hace innecesaria la utilización de un pegamento o soldadura líquida. La utilización de una técnica de este tipo es fundamental; de lo contrario, la columna capilar hueca podría resultar bloqueada por dicho líquido.

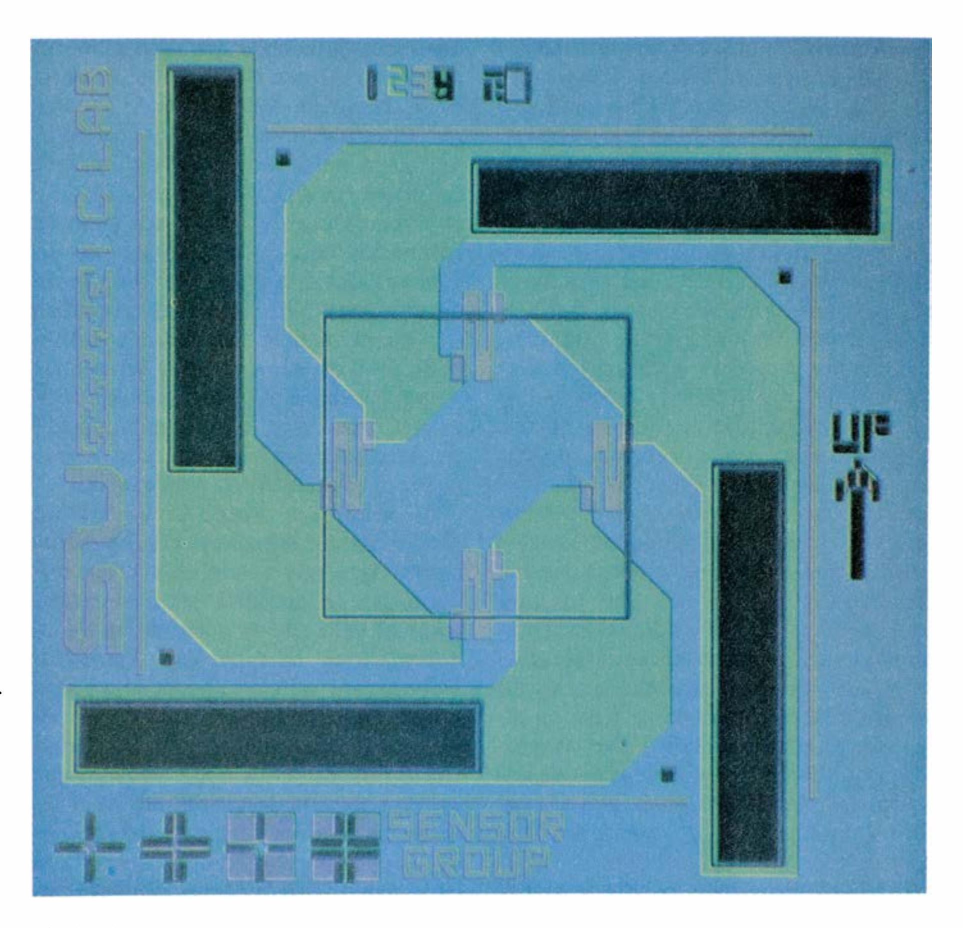
El proceso de unión anódica, denominado también unión electrostática o proceso Mallory, fue desarrollado por George Wallis y Daniel I. Pomerantz, de la firma P. R. Mallory & Co. La lámina de vidrio y la oblea de silicio se colocan en contacto y se calientan a unos 400 grados Celsius. Sobre la parte superior del vidrio se aplica un voltaje negativo elevado, en tanto que el silicio se conecta a tierra. El vidrio contiene una pequeña cantidad de sodio; a estas temperaturas el sodio se ioniza. Los iones positivos de sodio se mueven hacia el electrodo negativo, dejando tras de sí cargas negativas fijas en las proximidades del hueco existente entre el vidrio y el silicio. En virtud de lo cual, el voltaje a lo largo de dicho hueco es prácticamente igual al voltaje aplicado en el exterior. La atracción electrostática empuja al vidrio hacia el silicio, poniéndolos en contacto íntimo de forma tal que la unión que se crea es muy fuerte.

El detector de conductividad térmica, desarrollado por John H. Jerman,

11. DETECTOR DE CONDUCTIVIDAD TERMICA utilizado en los cromatógrafos de gases. Consta de un resistor construido mediante un proceso de grabado químico de una película metálica delgada depositada sobre una oblea de silicio recubierta por un vidrio. Una vez formado el resistor, se somete la oblea a un ataque anisotrópico por la parte posterior, tras el cual sólo queda una membrana delgada de vidrio bajo el resistor. Dado que el vidrio es un mal conductor térmico (a diferencia del silicio que es un buen conductor), la membrana aísla térmicamente al resistor. Cualquier cambio en la conductividad térmica del gas que fluye sobre el resistor alterará su temperatura y, en consecuencia, su resistencia eléctrica.

de Stanford, para el cromatógrafo de gases, constituye una unidad diferenciada soportada por una oblea de silicio independiente del resto del sistema. La oblea se oxida y se recubre con una película de vidrio pyrex depositada con la ayuda de una técnica especial ("chisporroteo"). El proceso se inicia colocando unos resistores de lámina delgada en la parte superior del vidrio. A continuación, se elimina todo el silicio con la ayuda de un grabador anisotrópico que actúa a través de ciertas aberturas cuadradas practicadas en el óxido que recubre la parte posterior de la oblea de silicio; procediendo de este modo se obtiene una membrana de vidrio térmicamente aislante bajo cada conjunto de resistores. Más tarde, la oblea es seccionada en elementos, a los que se conectan sendos conductores. Finalmente, la superficie frontal de uno de esos elementos se conecta al canal de gas situado al final de la columna capilar.

Dado que la columna capilar tiene un volumen muy pequeño, la cantidad de gas de transporte que requiere el cromatógrafo de gases en miniatura es mucho menor que la de un cromatógrafo convencional. Este dato y el pequeño tamaño de la propia unidad permiten construir instrumentos de cromatografía de gases portátiles. De hecho, la empresa Microsensor Technology está desarrollando un instrumento portátil constituido por cinco cromatógrafos miniaturizados y un microordenador. Si las cinco columnas capilares están recubiertas por materiales diferentes, este instrumento puede ser sensible a unos 100 gases diferentes. El instrumento al que nos estamos refiriendo puede controlar de forma continua la calidad del aire o medir el contenido energético del gas que circula por un conducto de gas natural.



12. SENSOR DE PRESION DE SILICIO, representado aquí en una fase intermedia de su proceso de construcción. El cuadrado que aparece en el centro de la microfotografía es un diafragma delgado, sensible a la presión, de 0,5 milímetros de lado. Los elementos en forma de cinta de color violeta son resistores obtenidos por difusión de boro en la oblea de silicio. Las regiones de color verde tienen una concentración de boro elevada y actúan de conductores entre los resistores. Los rectángulos negros situados cerca de los bordes del sensor son surcos en forma de *V* que posteriormente se rellenarán de oro al que se pueden conectar los oportunos conductores. Cuando el diafragma se flexiona, la resistencia de las zonas violeta se altera por la tensión mecánica y en razón del efecto de piezorresistencia. Los resistores están situados de modo que la resistencia de dos de ellos aumente y la de los otros dos disminuya. La sensibilidad del resistor a los cambios de presión es así el doble de la que se obtendría con uno sencillo.

La micromecanización se ha utilizado recientemente en la fabricación de un sumidero de calor de elevado rendimiento en una oblea de silicio. Estos sumideros de calor se producen con el fin de mejorar el rendimiento de los circuitos integrados de muy alta velocidad. En general, la disipación de calor en un dispositivo electrónico aumenta a medida que lo hace su velocidad, de forma tal que el calor liberado en los circuitos muy rápidos podría llegar a dificultar su funcionamiento si no se enfriaran convenientemente. En los sistemas de empaquetado especial, propios de los circuitos de muy alta velocidad, el límite superior de disipación de potencia se sitúa en los 20 watt por centímetro cuadrado de lámina.

David B. Tuckerman y Fabian Pease, del Laboratorio de Circuitos Integrados de Stanford, han desarrollado un sumidero de calor capaz de dar cuenta de potencias mucho más elevadas. La idea consiste en fabricar los dispositivos electrónicos de alta velocidad sobre la superficie delantera de una oblea de silicio y grabar una serie de surcos rectangulares poco espaciados entre sí en la parte trasera. Por estos canales circula un líquido de refrigeración. La conductividad térmica del propio silicio es muy alta y Tuckerman y Pease han fabricado dispositivos en los que se ha variado el espesor de las paredes, la anchura de los canales y la profundidad de los canales con el fin de obtener un eficaz enfriamiento.

Los canales rectangulares se producen de forma análoga a la utilizada en el caso de los canales en forma de *V*, descritos antes. La acción de un grabador anisotrópico, que interviene a través de una figura de óxido consistente en una serie de líneas paralelas delgadas, crea otros tantos canales en forma de *V* si la oblea es del tipo <100>. Ahora bien, si la oblea es del tipo <110>, la acción de un grabador anisotrópico que intervenga a través de la misma figura de óxido a la que nos acabamos de referir puede crear canales de paredes verticales cuya sección transversal es prácticamente rectangular. En los sumideros de calor, se prefieren los canales rectangulares; dichos canales pueden estar menos espaciados que los canales en *V* y, por tanto, la superficie expuesta al líquido de enfriamiento resulta ser mayor. Una vez que los canales han sido grabados, la oblea se une anódicamente a un vidrio con el fin de crear una serie de microconductos internos. Un sumidero de calor dotado de una serie de canales de 50 micrometros de anchura, 300 micrometros de profundidad y separados entre sí 100

micrometros, y enfriado por circulación forzada de agua es capaz de disipar más de 1000 watt por centímetro cuadrado de lámina.

Para construir estos sumideros de calor, hay que recurrir a obleas de silicio con una determinada orientación. Tuckerman y Pease han investigado también la posibilidad de crear sumideros de calor grabando los canales por medios mecánicos en lugar de químicos. Utilizando una hoja de sierra circular recubierta con polvo de diamante se pueden obtener canales de 50 a 200 micrometros de anchura sobre la superficie de una oblea de silicio. Aunque las dimensiones de los canales se controlan con mayor dificultad si utilizamos estos medios mecánicos, con ellos se pueden obtener sumideros de calor usando obleas de silicio de cualquier orientación y obleas de otros materiales, arseniuro de galio por ejemplo, que es un semiconductor utilizado de base en algunos circuitos integrados. Por otro lado, la grabación de las obleas puede adquirir forma de entramado, con lo cual estos circuitos están soportados por una especie de microcolumnas; utilizando esta disposición es menos probable que se produzca una obstrucción en el sumidero de calor motivada por la presencia de partículas extrañas en el líquido de enfriamiento.

Los acelerómetros y sensores de presión son otros tantos ejemplos de sensores de silicio integrados. El acelerómetro más pequeño fabricado hasta ahora es un dispositivo de puente de óxido provisto de una unidad de amplificación. El acelerómetro, diseñado por Petersen, Anne C. Shartel y Norman F. Raley, del Laboratorio de Investigación San José de la IBM, consiste en una cinta de óxido suspendida sobre un pozo hueco obtenido mediante la técnica de detención de la grabación con boro. El proceso se inicia depositando una capa metálica sobre la superficie del puente de óxido. La capa de metal y el silicio plano del fondo del pozo se comportan como las dos placas de un condensador con dieléctrico de aire de espesor variable. En el extremo libre del puente se sitúa una pieza de oro. Si la lámina de oro se mueve repentinamente la inercia del peso de oro produce la flexión de la cinta, lo cual modifica el espesor del hueco de aire y, por ende, la capacidad del condensador.

La salida del sensor es un voltaje proporcional a la aceleración. El instrumento tiene una sensibilidad de dos milivolt por *g* (unidad de aceleración igual al valor de la aceleración de la gravedad en la superficie terrestre). Deberíamos señalar que el amplificador es una parte fundamental del dispositivo. El cambio en la capacidad susceptible de ser detectado es muy pequeño y pasaría desapercibido en el seno de los cambios de capacidad relacionados con los cables unidos al sensor si no se amplificara la correspondiente señal. En general, con sensores pequeños se impone acondicionar la señal antes de proceder a su transmisión.

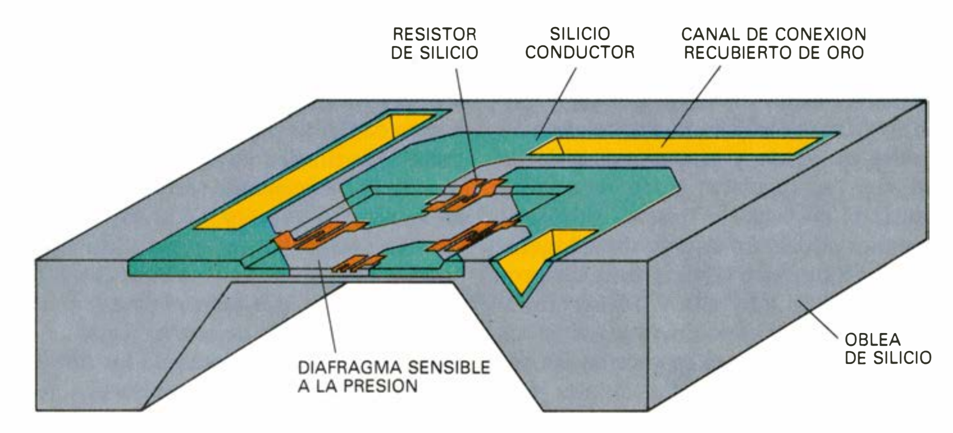
Los sensores de silicio más eficaces con que contamos en la actualidad son los sensores de presión, algunos de los cuales incorporan dispositivos electrónicos integrados. Hay dos tipos fundamentales de sensores de presión de silicio: sensores piezorresistivos y capacitativos. La resistencia de los resistores eléctricos situados en un diafragma delgado flexible de un sensor piezorresistivo cambia al flexionar dicho diafragma. En un sensor capacitativo el diafragma sensible es una de las placas de un condensador con dieléctrico de aire de espesor variable.

En la mayoría de los casos el diafragma de un sensor de presión está grabado anisotrópicamente en una oblea de silicio de tipo <100>. Controlando el tiempo de grabado se obtienen diafragmas con espesores comprendidos entre 10 y 50 micrometros. Se consiguen diafragmas de espesores todavía menores recurriendo a técnicas de detención del proceso de grabado.

Los resistores de un sensor piezorresistivo se obtienen contaminando ciertas zonas del diafragma. Existen dos tipos de contaminantes. Los elementos situados a la derecha del silicio en el sistema periódico, tales como el fósforo, tienen cinco electrones de valencia, en tanto que el silicio posee cuatro. El electrón adicional puede convertirse en un portador de corriente eléctrica. Los elementos situados a la izquierda del silicio en el sistema periódico, tales como el boro, tienen tres electrones de valencia. La ausencia de un electrón (es decir, la presencia de un hueco) puede actuar también como un portador de corriente, aunque los huecos son menos móviles que los electrones. La corriente no fluirá desde una región cargada con fósforo hasta una región cargada con boro, debido a la diferencia en los portadores de carga. En consecuencia, si el silicio que forma la oblea base estuviera ligeramente contaminado se podría difundir una concentración mayor de boro en ciertas áreas seleccionadas de la oblea con el fin de obtener resistores eléctricamente aislados del resto del diafragma.

El efecto piezorresistivo consiste en el cambio de la resistencia eléctrica de un material como consecuencia de una tensión mecánica; la magnitud del efecto depende de la orientación de los resistores. Por una afortunada coincidencia, los resistores cargados con boro en una oblea <100> se muestran más sensibles a las tensiones cuando están alineados con los extremos <110> de una cavidad grabada anisotrópicamente.

Nosotros hemos desarrollado un sensor de presión piezorresistivo con cuatro resistores sobre un diafragma cuadrado. Los resistores se encuentran cerca de los bordes del diafragma (las zonas de máxima tensión) y están alineados con las direcciones <110> (las direcciones para las que la respuesta a las tensiones es máxima). Los resistores están unidos a una configuración de dos



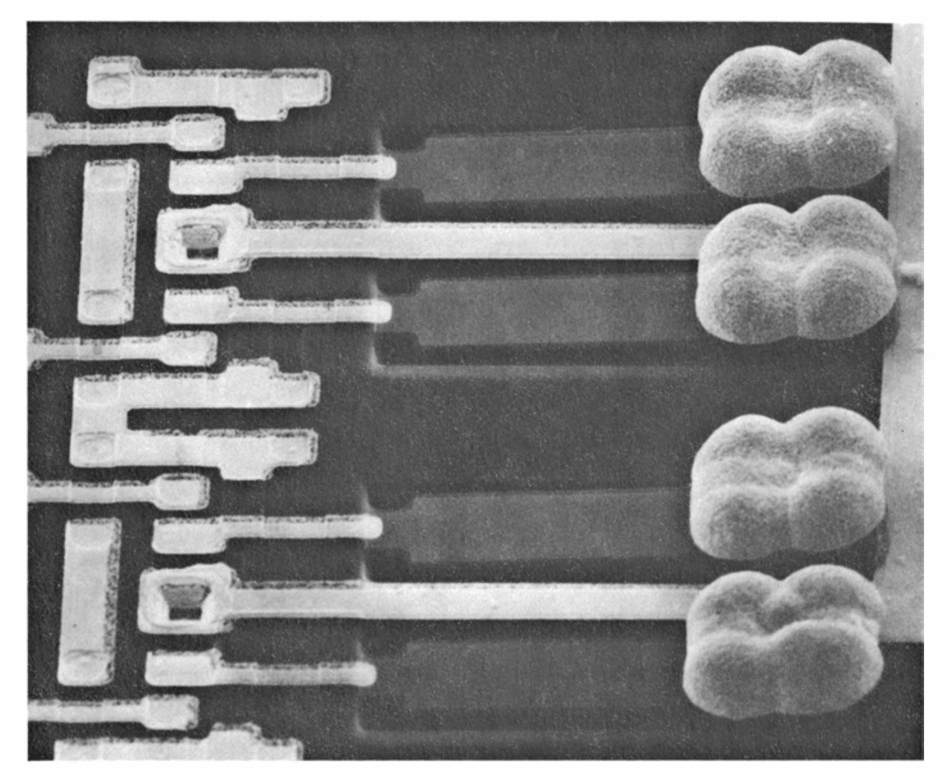
13. SECCION TRANSVERSAL de un sensor de presión en la que el espesor del diafragma se ha exagerado un tanto con el fin de que se aprecien todos los detalles. El diafragma se define mediante un grabado anisotrópico poco profundo en la parte superior de la oblea y un grabado mucho más enérgico en la parte inferior. Terminada la fabricación del sensor, la superficie se une a una lámina de vidrio, de suerte que se cree allí una cavidad cerrada. El fondo del diafragma se expone a la presión exterior a medir. Se conoce la presión en la cavidad; la señal de salida proporciona, pues, una medida absoluta de dicha presión.

terminales conocida por puente de Wheatstone. Cuando el diafragma se flexiona, la resistencia de dos de los resistores aumenta y la de los otros dos disminuye en igual cuantía. Como consecuencia, la sensibilidad del puente a los cambios de presión es el doble de lo que sería en un único resistor.

La primera unidad de sensor de presión con circuito activo (es decir, un circuito con transistores que amplifican la señal) fue diseñada por John M. Borky y Kensall Wise, de la Universidad de Michigan. Este circuito convierte el voltaje de salida de un puente de Wheatstone en una señal de frecuencia modulada (FM); en otras palabras, los cambios de presión se traducen en cambios de frecuencia. Si se asigna un intervalo de frecuencias diferente a cada sensor, dichos sensores pueden compartir una línea de transmisión de forma análoga a como las distintas emisoras de radio de FM comparten la banda de radiofrecuencias.

Los sensores de presión capacitativos son más sensibles a los cambios de presión que los sensores piezorresistivos, pero, tal como señalamos anteriormente, la señal de un sensor capacitivo se puede perder fácilmente a causa del ruido. Los circuitos de acondicionamiento de la señal mejoran el rendimiento de los sensores piezorresistivos; ahora bien, en el caso de los sensores capacitativos, la importancia de esta mejora es crucial. Craig S. Sander, James W. Knutti y James D. Meindl, del Laboratorio de Circuitos Integrados de Stanford, diseñaron el primer sensor de presión capacitativo con una unidad de circuito activo. El diafragma de este sensor está situado a una profundidad de unos cinco micrometros en la oblea de silicio y actúa como una placa de condensador. La otra placa está constituida por una lámina metálica unida a una lámina de vidrio que, a su vez, está unida anódicamente a la lámina de silicio.

Una cantidad fija de gas está encerrada en la cavidad formada por el diafragma y la lámina de vidrio recubierta por el metal. El otro lado del diafragma está expuesto a la presión que tratamos de medir. Cuando la presión cambia y el diafragma se flexiona, el espesor de la capa de aire situado entre el metal y el silicio cambia. El circuito activo controla las modificaciones resultantes en la capacidad del condensador, produciendo una serie de impulsos de voltaje. El tiempo transcurrido entre dos de estos impulsos es proporcional a la presión que tratamos de medir. Esta señal de impulsos codificados en el tiempo, al igual que la señal de FM que procede de los sensores de presión piezorresistivos, es menos sensible al ocultamiento por ruido de lo que lo es una simple señal de voltaje sin codificar.



14. ACELEROMETROS CAPACITIVOS. Son puentes voladizos de dióxido de silicio tendidos sobre ciertas cavidades practicadas en el silicio. En esta microfotografía electrónica se muestra un conjunto de cuatro acelerómetros. Las piezas de oro que se pueden observar en el extremo de la derecha de cada uno de los puentes aumentan la inercia de estos dispositivos y, por tanto, su sensibilidad. La capa metálica de la parte superior del puente, la capa del fondo del pozo, constituida por boro conductor como elemento de detención del grabado, y el aire interpuesto entre ambos constituyen los elementos de un condensador variable. Cuando se le somete a una aceleración, el puente se flexiona y la capacidad de dicho condensador se altera. Los circuitos que aparecen a la izquierda de la figura amplifican la correspondiente señal; la señal de salida del sensor es una diferencia de potencial que resulta proporcional a la aceleración.

Los sensores integrados pueden satisfacer las exigencias de un mayor número de aplicaciones que los sensores discretos, ya que las señales son relativamente insensibles al ruido y a las pérdidas. En consecuencia, los sensores integrados se pueden utilizar en ambientes difíciles, tales como el que existe bajo la carrocería de un automóvil, o en situaciones en las que el dispositivo sensible ha de estar separado físicamente de la electrónica que procesa la señal, como sucede en ciertas aplicaciones de biomedicina. Muchos de los sensores que se encuentran en fase de desarrollo se están diseñando con el fin de poder afrontar con éxito estas condiciones adversas. Barth y Angell, coautores de este artículo, están trabajando en sensores susceptibles de ser implantados en tejidos vivos.

También se están desarrollando ciertas unidades elementales provistas de diferentes tipos de sensores (de temperatura y presión), varias copias de un mismo sensor (por motivos de fiabilidad) y unidades que incluyen la posibilidad de recalibrar los sensores. En el futuro, muchos de los sensores integrados estarán capacitados para convertir señales analógicas en señales digitales, de forma tal que sea posible su conexión directa a circuitos electrónicos digitales. En general, uno de los mayores retos con que se enfrenta esta técnica consiste en determinar la cantidad de circuitos que se pueden incorporar en una misma unidad sensora, desde el punto de vista físico y desde el punto de vista del coste económico.

Al mismo tiempo que se dedica gran atención al tema de los sensores, se siguen desarrollando los dispositivos de silicio estrictamente mecánicos. Las técnicas de grabado de silicio se encuentran en fase de mejora y, sin duda alguna, en los próximos años emergerán de los laboratorios nuevas estructuras micromecánicas. Tal vez llegue un día en el que los dispositivos mecánicos de silicio sean algo tan corriente como lo son en la actualidad los dispositivos electrónicos de silicio.

Estructura de quarks y leptones

Considerados las partículas últimas de la materia, parece, sin embargo, que pudieran estar formados por otras entidades, confinadas en un volumen menor que la milésima parte del tamaño de un protón

Haim Harari

En los últimos cien años, la búsqueda de los constituyentes fundamentales de la materia ha profundizado en cuatro niveles estructurales. Se demostró que toda la materia constaba de átomos. Se puso de manifiesto que el propio átomo era un núcleo denso rodeado por una nube de electrones. A su vez, el núcleo se descompuso en protones y neutrones. Más cerca en el tiempo, se averiguó que protón y neutrón eran también partículas compuestas; están formados por entidades menores llamadas quarks. ¿Qué vendrá luego? Es muy posible que el descubrimiento de orbes dentro de otros orbes haya alcanzado su final definitivo y que no tengamos forma de dividir los quarks. Por su parte, los leptones, el tipo de partículas que incluye al electrón, podrían ser también elementales e indivisibles. Algunos físicos, sin embargo, no están del todo seguros que el corazón más interno de la materia se haya desentrañado. Y se empiezan a preguntar si los quarks y leptones no poseerán, a su vez, cierta composición interna.

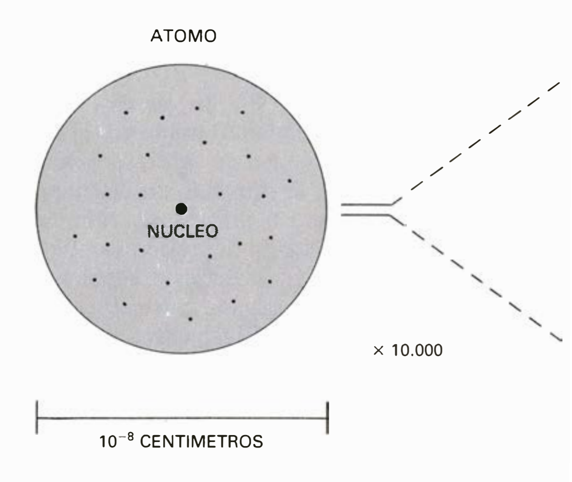
Induce a considerar otro nivel de estructura más la convicción (¿prejuicio quizá?) de que tendría que haber sólo unos pocos bloques fundamentales a partir de los cuales se construiría la materia. La historia nos atestigua que la economía de medios ha sido durante mucho tiempo un principio guía de la física, rindiendo buenos servicios hasta hoy. La lista de los constituyentes fundamentales de la materia primero creció inverosímilmente hasta finales del siglo XIX, cuando el número de elementos químicos, y por tanto el número de especies de átomos, se acercaba al centenar. La resolución de la estructura atómica acabó con el problema y, alrededor de 1935, el número de partículas elementales se había rebajado a cuatro: protón, neutrón, electrón y neutrino. Esta sobria visión del mundo entró en crisis en los años 50 y 60 cuando comenzó a verse que el protón y el neutrón pertenecían a una familia muy nutrida de partículas, a la que hemos apellidado la familia de los hadrones. A mitad de los 60, el número de formas fundamentales de la materia volvía a rozar el centenar. Llegó entonces el modelo quark para aliviar la situación. En la formulación inicial del mismo, todos los hadrones se explicaban por combinaciones de sólo tres tipos de quarks.

Ahora, ante la proliferación de quarks y leptones, se aviva el interés por hallar un esquema más sencillo. El modelo original hablaba de tres quarks; hoy se cifra su número mínimo en 18, amén de seis leptones y otra docena de partículas que actúan como portadores de las fuerzas. Tres docenas de unidades básicas de la materia son demasiadas para la sensibilidad de algunos físicos, y nada hace pensar que no se descubran nuevos quarks y leptones. Postular un nivel más profundo de organización es, quizás, el camino más expedito para reducir la lista. Todos los quarks y leptones serían, así, objetos compuestos, como lo son los átomos y hadrones, y deberían su variedad al número de maneras distintas en que pueden combinarse unos pocos constituyentes más pequeños. La diversidad que observamos de la naturaleza sería combinatoria, no intrínseca.

Vaya por delante que no disponemos todavía de señal alguna de ningún tipo de estructura interna, ni para quarks ni para leptones. En el caso de los leptones, los experimentos han sondeado hasta 10^{-16} centímetros sin encontrar nada que contradiga la presunción según la cual los leptones son puntuales y carecen de estructura. Para el quark, al no ser posible examinarlo aislado, la falta de datos sobre cualquier característica interna es mucho mayor. No terminan aquí las dificultades. Desde una perspectiva estrictamente teórica, cuesta fundamentar la idea de las subpartículas: nadie ha sido capaz de trazar una explicación coherente de cómo éstas se moverían dentro de un quark o un leptón o cómo interaccionarían mutuamente. Serían sutilísimas. Si un átomo viera aumentado su tamaño hasta el de la Tierra, sus constituyentes más internos no superarían el de una naranja. Pero ello no impide que los modelos que introducen una subestructura para quarks y leptones atraigan poderosamente al sentido estético y a la imaginación: levantan un mundo complejo con un número limitado de materiales.

Toda teoría sobre las partículas elementales de la materia ha de considerar las fuerzas que actúan entre aquéllas y las leyes de la naturaleza que gobiernan las fuerzas. Menguada sería la ganancia si simplificáramos el espectro de partículas a costa de elevar el número de fuerzas y leyes. A lo largo de la historia de la física vemos un sutil entrelazamiento entre la lista de partículas y la lista de fuerzas.

Alrededor de 1800, cuatro fuerzas se consideraban fundamentales: la gravitación, la electricidad, el magnetismo y la fuerza de corto alcance entre moléculas, responsable de la cohesión de la



1. LA JERARQUIA DE PARTICULAS en la estructura de la materia se escalona hoy en cuatro niveles. Toda la materia está hecha de átomos; el

materia. Una sucesión de notables descubrimientos experimentales y teóricos condujeron luego al reconocimiento de que electricidad y magnetismo eran, en realidad, dos manifestaciones de una misma fuerza básica, que pronto recibiría el nombre de electromagnetismo. El descubrimiento de la estructura atómica trajo una nueva revisión. Aunque un átomo es, en conjunto, eléctricamente neutro, sus constituyentes están cargados; la fuerza molecular de corto alcance debía entenderse, pues, como un complicado efecto residual de las interacciones electromagnéticas entre los núcleos positivos y los electrones negativos. Cuando dos átomos neutros están alejados, no median prácticamente fuerzas electromagnéticas entre ellos. Pero cuando se hallan cerca uno del otro, los constituyentes cargados de un átomo son capaces de "ver" e influir en las cargas internas del otro, dando lugar a diversas atracciones y repulsiones de corto alcance.

Merced a esos avances, la física se quedó con sólo dos fuerzas básicas. La unificación de la electricidad y el magnetismo redujo su número en una unidad, mientras la interacción molecular fue relegada del rango de fuerza fundamental al de fuerza derivada. Las dos fuerzas fundamentales que subsistían, gravitación y electromagnetismo, eran ambas de largo alcance. Sin embargo, la exploración de la estructura nuclear introdujo pronto dos nuevas fuerzas de corto alcance. La fuerza fuerte enlazaba protones y neutrones juntos en el núcleo y la fuerza débil mediaba ciertas transformaciones de una partícula en otra, como en la desintegración beta de un núcleo radiactivo. Volvíamos a tener cuatro fuerzas.

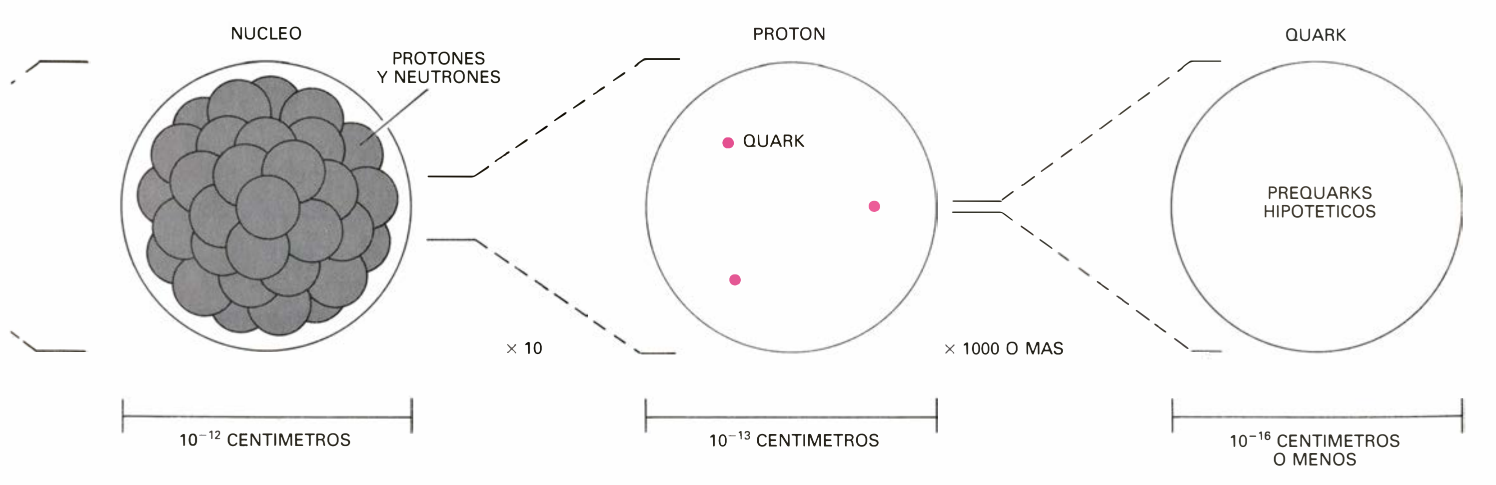
El desarrollo del modelo quark y de la correspondiente teoría de las interacciones entre quarks presentó la ocasión para revisar la lista de fuerzas. Los quarks de un protón o un neutrón deben considerarse unidos por una nueva fuerza fundamental de largo alcance llamada fuerza de color, que actúa sobre los quarks porque éstos llevan un nuevo tipo de carga, el color (ni la fuerza ni la carga tienen relación alguna con los colores del espectro). Así como un átomo lo integran constituyentes dotados de carga eléctrica, aunque en cuanto tal sea neutro, el protón o el neutrón constan de quarks coloreados, pero en cuanto tales carecen de color. Cuando dos protones, carentes de color, están alejados no hay de hecho fuerzas de color entre ellos; ahora bien, al aproximarse los quarks coloreados de un protón "ven" las cargas de color del otro protón. Las atracciones y repulsiones de corto alcance que así se generan se han identificado con los efectos de la fuerza fuerte. En otras palabras: la fuerza molecular de corto alcance se convirtió en residuo de la fuerza electromagnética, de largo alcance; la fuerza fuerte de corto alcance, en un residuo de la fuerza de color, también de largo alcance.

Antes de cerrar este breve resumen de la historia de las fuerzas de la naturaleza, conviene agregar un último capítulo. Se ha encontrado una profunda relación, no carente de belleza, entre el electromagnetismo y la fuerza débil que las catapulta hasta las mismas puertas de la total unificación. A pesar de su obvia relación, la conexión no es tan estrecha como en el caso de electricidad y magnetismo; de momento habrá que seguir suponiéndolas fuerzas distintas. Por consiguiente, la lista de fuerzas fundamentales que se maneja en física consta todavía de cuatro términos: la fuerza gravitacional, la electromagnética y la de color, las tres de largo alcance, y la fuerza débil, de corto alcance. Dentro de los límites de nuestro conocimiento, todos los fenómenos de la naturaleza se interpretan a través de esas fuerzas y de sus efectos residuales.

La evolución de las ideas sobre las partículas y la correspondiente a las ideas sobre las fuerzas son interdependientes. En cuanto se registran nuevas partículas fundamentales, las antiguas se convierten en objetos compuestos. En cuanto se encuentran nuevas fuerzas, las antiguas se unifican o quedan reducidas a un status residual. Las listas de partículas y fuerzas se revisan a medida que se explora la materia a escala menor y avanza el conocimiento teórico. Cualquier cambio en una lista lleva inevitablemente a una modificación en la otra. Las recientes hipótesis sobre la estructura de quarks y leptones no constituyen ninguna excepción; piden también cambios en la asignación de fuerzas. Queda aún por ver si tales cambios representan una simplificación.

De las cuatro fuerzas fundamentales establecidas, la gravitación ocupa una categoría aparte. Demasiado débil para ser detectada en las interacciones entre partículas individuales, no puede explicarse en términos de sucesos microscópicos. Para las otras tres fuerzas se han desarrollado con éxito teorías que hoy gozan de amplia aceptación. Aunque distintas, las tres teorías muestran una coherencia mutua; si las tomamos en conjunto, constituyen un modelo general de las partículas elementales y sus interacciones. Aquí lo denominaremos modelo standard.

En el modelo standard los constituyentes indivisibles de la materia son los quarks y los leptones. Empezare-

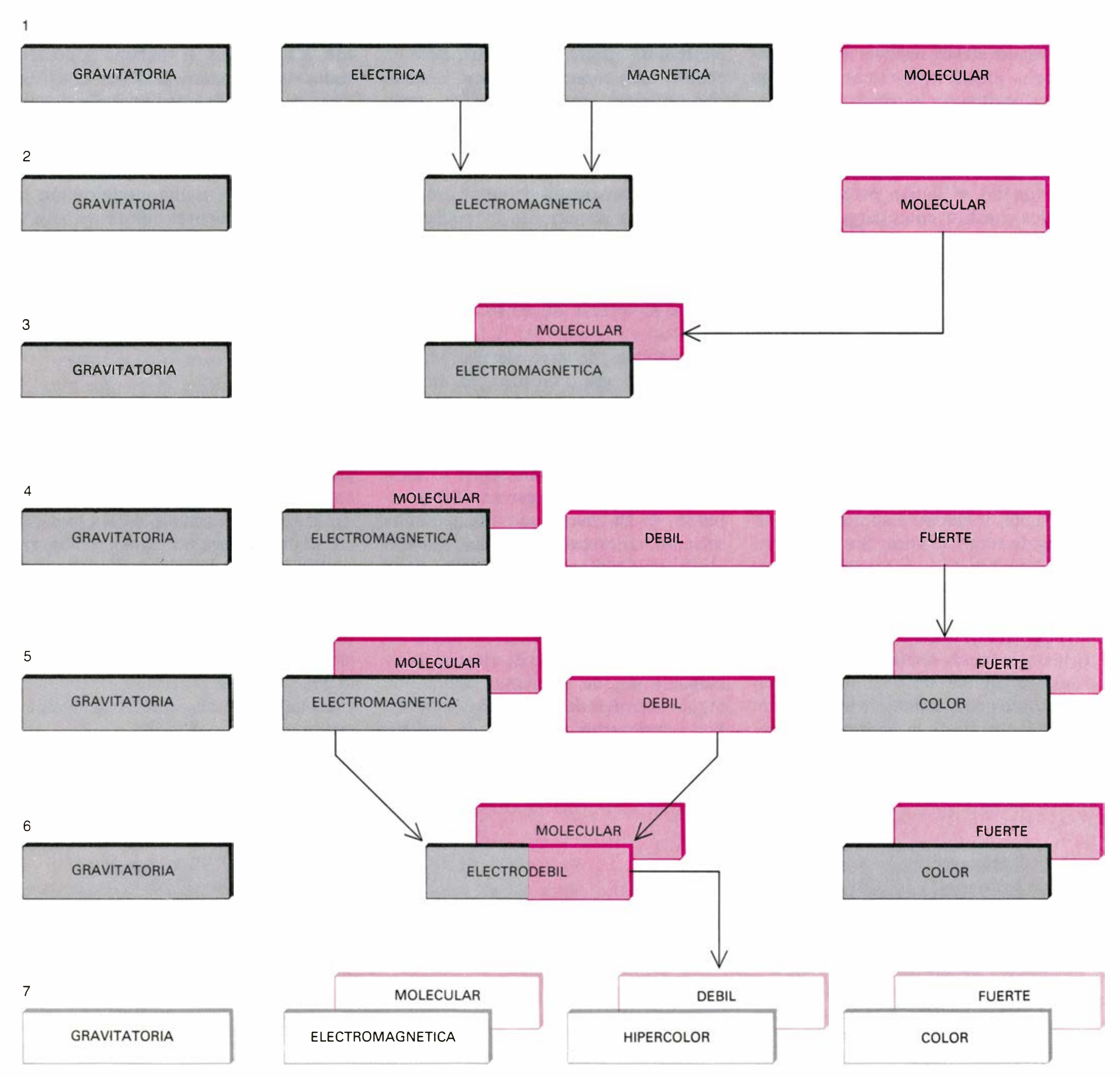


átomo consta de un núcleo rodeado de electrones; el núcleo se compone de protones y neutrones; cada protón y neutrón se cree formado por tres quarks. Los desarrollos teóricos recientes hablan de añadir un quinto nivel: el quark podría estar compuesto por hipotéticos constituyentes más sutiles, que llamaremos aquí, genéricamente, prequarks. Los leptones, la clase de partículas que incluye el electrón, podrían estar constituidos, también, por prequarks.

mos por estos últimos. Hay seis de ellos: el electrón y su compañero el neutrino electrónico; el muon y el neutrino muónico y la partícula tau y el neutrino tauónico. Electrón, muon y tau tienen carga eléctrica −1; los tres neutrinos son eléctricamente neutros.

Hay también seis tipos básicos de quark, que han recibido los nombres de "arriba" (*up*), "abajo" (*down*), "extraño" (*strange*), "encanto" (*charm*), "cima" (*top*) y "fondo" (*bottom*), o *u, d, c, s, t* y *b*. (El quark "cima" no se ha detectado aún experimentalmente, como tampoco lo ha sido el neutrino tauónico, pero pocos físicos teóricos dudan de su existencia.) Los quarks *u, c* y *t* tienen carga eléctrica + 2/3; los *d, s* y *b* carga − 1/3. Cada tipo de quark posee, además, tres posibles colores, que designaré por rojo, amarillo y azul. Así, pues, si cada quark coloreado se cuenta separadamente como una partícula habrá, en total, 18 variedades de los mismos. Debe señalarse que cada quark lleva tanto color como carga eléctrica; los leptones carecen de color.

Por cada partícula de este esquema existe una antipartícula con la misma masa pero valores opuestos de carga eléctrica y color. La antipartícula del electrón es el positrón, que tiene carga + 1. La antipartícula de un quark *u*



2. CLASIFICACION DE LAS FUERZAS FUNDAMENTALES de la naturaleza en un esquema desarrollado conjuntamente con la lista de las partículas elementales. Las fuerzas de largo alcance aparecen dibujadas en gris; las de corto alcance, en color. A principios del siglo XIX se suponían fundamentales tres fuerzas de largo alcance: la gravitatoria, la eléctrica y la magnética. Una fuerza de corto alcance, la fuerza molecular responsable de la cohesión de la materia, tenía también carácter de fundamental. James Clerk Maxwell unificó electricidad y magnetismo; con el descubrimiento de la estructura atómica se hizo evidente que la fuerza molecular no era fundamental, sino que se trataba de un efecto residual del electromagnetismo. El descubrimiento del núcleo atómico introdujo dos fuerzas de corto alcance, la débil y la fuerte. En el modelo quark, la fuerza fuerte se convierte en un residuo de una nueva fuerza de largo alcance, la llamada fuerza de color. Se ha descubierto una profunda relación entre la fuerza débil y el electromagnetismo, de manera que pueden considerarse parcialmente unificados. La sexta línea del dibujo representa las fuerzas de la naturaleza tal como se entienden hoy en día en el "modelo standard" de la física de partículas elementales. Un modelo fructífero para la estructura de quarks y leptones podría traer una revisión adicional. En algunos modelos, por ejemplo, hay una nueva fuerza de largo alcance llamada hipercolor; la fuerza débil es un residuo suyo. En modelos de este tipo, todas las fuerzas fundamentales de la naturaleza son de largo alcance. Pero se trata de constructos teóricos donde predominan las hipótesis.

rojo, con carga + 2/3, es un antiquark $\bar{u}$ antirrojo, con carga − 2/3.

El color de los quarks recuerda, en muchas cosas, la carga eléctrica, aunque, por mor de la existencia de tres posibles colores, se trata de una propiedad bastante más complicada. Las partículas dotadas de carga eléctrica se juntan y forman un sistema eléctricamente neutro de una sola manera: combinando la misma cantidad de cargas eléctricas positivas y negativas. La partícula compuesta carente de color se forma, a partir de los quarks coloreados, de un mismo modo: combinando un quark coloreado con un antiquark anticoloreado. Pero, en el caso del color, hay una segunda vía para constituir un estado neutro: todo sistema compuesto con la misma cantidad de los tres colores o de los tres anticolores carece también de color. Ahí reside la razón de que el protón, que consta de un quark rojo, otro amarillo y un tercero azul, no posea color.

No debemos olvidar otra propiedad de quarks y leptones: cada partícula tiene un espín, o momento angular intrínseco, igual a la mitad de la unidad mecánico-cuántica básica del momento angular. Cuando una partícula con espín 1/2 se mueve a lo largo de una línea recta, su rotación intrínseca puede seguir el sentido de las agujas del reloj o el sentido contrario, mirando la partícula en la dirección del movimiento. Si el espín procede en el sentido del reloj, a la partícula se le dice dextrógira ("right-handed"), porque, cuando los dedos de la mano derecha giran en la misma dirección del espín, el pulgar señala la dirección del movimiento. Para una partícula con espín en sentido contrario, la regla de la mano izquierda describe el movimiento, y entonces se dice que la partícula es levógira ("left-handed").

		QUARKS		LEPTONES	
GENERACIONES	PRIMERA	*u* (ARRIBA)	*d* (ABAJO)	*e* (ELECTRON)	ν_e (NEUTRINO ELECTRONICO)
	SEGUNDA	*c* (ENCANTO)	*s* (EXTRAÑO)	μ (MUON)	ν_μ (NEUTRINO MUONICO)
	TERCERA	*t* (CIMA)	*b* (FONDO)	τ (TAU)	ν_τ (NEUTRINO TAUONICO)
CARGAS	ELECTRICA	+2/3	−1/3	−1	0
	COLOR	ROJO / AMARILLO / AZUL	ROJO / AMARILLO / AZUL	SIN COLOR	SIN COLOR
FUERZAS	COLOR				
	ELECTRO-MAGNETICA				
	DEBIL				

3. **MODELO STANDARD de las partículas elementales. Incluye tres "generaciones" de quarks y leptones, aunque toda la materia ordinaria puede construirse con las partículas de la primera generación solamente. Los quarks se distinguen por el valor fraccional de su carga eléctrica y por una propiedad que recibe el nombre caprichoso de color: cada tipo de quark aparece en versiones roja, amarilla y azul. Los leptones tienen unidades enteras de carga eléctrica y carecen de color. Una y otra clase de partículas difieren también en su respuesta a las diversas fuerzas. Sólo los quarks se hallan sujetos a la fuerza de color y, como resultado, permanecen confinados dentro de partículas compuestas, como el protón.**

En el modelo standard, una misma estructura matemática da razón de las tres fuerzas que actúan sobre quarks y leptones. Se la conoce como teoría de campos invariante de aforo, o teoría de aforo simplemente (o de contraste, en inglés "gauge theory"). Cada fuerza se transmite de una partícula a otra mediante campos portadores, que, a su vez, se encarnan en partículas portadoras, o bosones de aforo.

La teoría de aforo de la fuerza electromagnética, llamada electrodinámica cuántica, abreviada QED ("*Q*uantum *E*lectro*D*ynamics"), es la primera y más sencilla de las tres teorías. La crearon, en los años 40, Richard P. Feynman, Julian S. Schwinger y Sin-Itiro Tomonaga. La QED describe las interacciones entre partículas cargadas eléctricamente, en particular, el electrón y el positrón. Hay un tipo de bosón de aforo para mediar las interacciones; se trata del fotón, el cuanto familiar de la radiación electromagnética; carece de masa y no posee carga eléctrica propia. La QED constituye, quizá, la teoría física que ha recibido más ajustada verificación. Predice, por ejemplo, el momento magnético del electrón con una precisión de 10 cifras significativas, por lo menos.

A imagen de la QED, se formuló la teoría de la fuerza de color, conocida por cromodinámica cuántica, QCD ("*Q*uantum *C*hromo*D*ynamics"). En su desarrollo, a lo largo de casi veinte años, han puesto su empeño muchos físicos teóricos. En la QCD las partículas interaccionan en virtud de su color, y no en razón de su carga eléctrica. Los bosones de aforo de la QCD, responsables del enlace de los quarks dentro de un hadrón, se llaman gluones. Como el fotón, los gluones carecen de masa; ahora bien, en tanto que hay un solo tipo de fotón, existen ocho especies de gluon. Otra diferencia entre el fotón y los gluones adquiere una importancia mayor. Aunque el fotón es el intermediario de la fuerza electromagnética, no tiene carga eléctrica ni ocasiona, por tanto, fuerzas electromagnéticas sobre sí mismo (o, por lo menos, no de magnitud importante). Los gluones, por contra, sí poseen color. Transmiten la fuerza de color entre los quarks, pero tienen su propio color y responden a la fuerza de color. Esta reflexividad, en cuya virtud el portador de la fuerza actúa sobre sí mismo, hace sumamente difícil un análisis matemático completo de la fuerza de color.

Una peculiaridad que parece inherente a la QCD es el fenómeno del confinamiento del color. Se sabe que la fuerza de color encierra los objetos coloreados (quarks y gluones) dentro de objetos compuestos que, invariablemente, carecen de color (protones y neutrones). Las partículas coloreadas no pueden escapar jamás (a no ser que

formen nuevas combinaciones sin color). A causa del confinamiento del color, los físicos así lo suponen, nunca se ha visto aislado un quark o un gluon. Téngase en cuenta que, si bien la idea del confinamiento del color goza hoy de amplio favor, no se ha probado que se deduzca de la QCD. El futuro puede reservarnos alguna sorpresa.

La fuerza débil difiere un tanto de las otras dos, pero puede describirse también por una teoría de aforo del mismo tipo general. En los años 60 y principios de los 70, los físicos desarrollaron la teoría y establecieron la relación entre la fuerza débil y la electromagnética. En ese punto destacaron (en orden cronológico) Sheldon Lee Glashow, de la Universidad de Harvard, Steven Weinberg, de la Universidad de Texas en Austin, Abdus Salam, del Centro de Física Teórica de Trieste, y Gerard 't Hooft, de la Universidad de Utrecht.

Curiosamente, las cargas sobre las cuales actúa la fuerza débil están asociadas con la helicidad de la partícula. Tanto entre los quarks como entre los leptones, las partículas levógiras y las antipartículas dextrógiras tienen carga débil, en tanto que las partículas dextrógiras y las antipartículas levógiras son neutras respecto a la fuerza débil. Lo que resulta más extraño todavía es que la carga débil no se conserve en la naturaleza: una unidad de carga puede crearse de la nada o desaparecer en el vacío. En cambio, la cantidad neta de carga eléctrica en un sistema de partículas aislado no puede alterarse nunca, ni tampoco el color neto. La fuerza débil se distingue, asimismo, por su rango, cortísimo; sus efectos se extienden a 10^{-16} centímetros de distancia, la milésima parte del diámetro del protón.

En la teoría de aforo de la fuerza débil, la incapacidad de la carga débil por conservarse y el corto alcance de la fuerza se atribuyen a la ruptura espontánea de la simetría, mecanismo al que me referiré con cierto detalle más adelante. Baste señalar de momento que el mecanismo de ruptura de la simetría implica que la carga débil, y la helicidad asociada de las partículas, debería conservarse a energías extremadamente altas, cuando la masa de una partícula es una parte despreciable de su energía cinética.

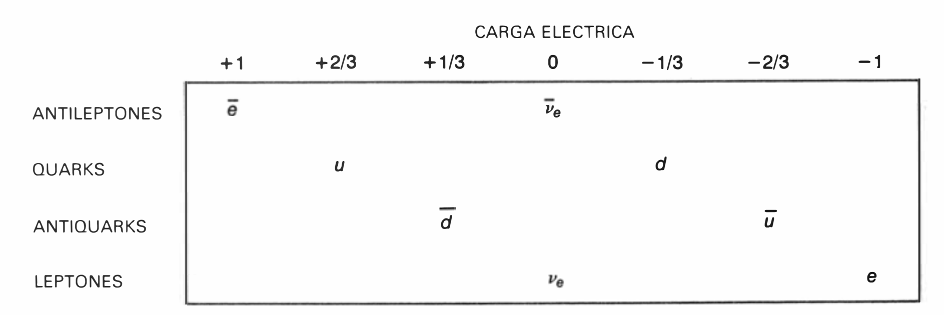
La ruptura espontánea de la simetría requiere también que los bosones de aforo de la fuerza débil sean partículas con masa; la verdad es que centuplican la del protón. En el modelo convencional hay tres bosones de ésos: dos de ellos, designados por W^+ y W^-, llevan tanto carga eléctrica como carga débil; el tercero, llamado Z^o, es eléctricamente neutro. La gran masa de los bosones débiles da cuenta del corto alcance de la fuerza. Siguiendo el principio de incertidumbre de la mecánica cuántica, el alcance de una fuerza resulta inversamente proporcional a la masa de la partícula que la transmite. Así, el electromagnetismo y la fuerza de color, transmitidos por bosones de aforo sin masa, poseen un alcance infinito, en tanto que la fuerza débil tiene una esfera de influencia sumamente pequeña. De la ruptura espontánea de la simetría se deriva otra consecuencia: predice la existencia de una partícula adicional con masa, por lo menos, distinta de los bosones débiles. Se le llama partícula de Higgs, en honor de Peter Higgs, de la Universidad de Edinburgh, quien contribuyó de manera importante a la teoría de la ruptura espontánea de simetría.

Los éxitos del modelo standard registrados a lo largo de los últimos diez años han constituido un gratificante espaldarazo para la labor de los físicos. Todas las formas conocidas de materia pueden construirse con los 18 quarks coloreados y los seis leptones del modelo. Todas las interacciones observadas de la materia pueden explicarse como intercambios de los 12 bosones de aforo incluidos en el modelo: el fotón, los ocho gluones y los tres bosones débiles. El modelo ofrece coherencia interna; ninguna parte está en contradicción con otra, y para todas las cantidades medibles se ha predicho un valor plausible y finito. La coherencia interna no es ninguna fruslería en un sistema conceptual que cubre un ámbito tan amplio. Hasta el momento, el modelo se muestra acorde también con todos los resultados experimentales; es decir, ninguna predicción clara del modelo ha resultado negada por los experimentos. Para decir toda la verdad, quedan varias predicciones importantes por verificar; en particular, falta por descubrir el neutrino tauónico, el quark "cima", los bosones débiles y la partícula de Higgs. La primera prueba directa de los bosones W fue recientemente comunicada por un grupo de físicos experimentales del CERN, laboratorio europeo para la física de partículas con sede en Ginebra. En los próximos años nuevos aceleradores de partículas y aparatos detectores más sensibles verificarán las predicciones del modelo aún pendientes. La mayoría de los físicos confían plenamente en su confirmación.

Si el modelo convencional se ha mostrado tan acertado, ¿por qué plantearse teorías más refinadas? No debe buscarse el motivo principal en una sospecha de que el modelo convencional esté equivocado, sino en la sensación de cierta limitación que produce. Aunque responde de cuantas preguntas da pie, son muchas las preguntas que quedan abiertas y muchas regularidades de la naturaleza siguen pareciendo arbitrarias o puras coincidencias. En una palabra, es el propio modelo el que espera aclaración.

El indicio más fuerte de algún principio organizativo más allá del modelo standard nos lo ofrece la proliferación de partículas elementales. Las propiedades conocidas de la materia no son tan numerosas o diversas que necesiten 24 partículas para representarlas. Es más, parecen abundar las repeticiones en el espectro de quarks y leptones. Hay tres leptones con carga eléctrica − 1, tres leptones neutros, tres quarks con carga + 2/3 y tres quarks con carga −1/3. Todo va por triplicado, sin ninguna razón aparente. Un mundo que se levantara escogiendo una partícula de cada uno de los cuatro grupos aportaría toda la variedad necesaria.

Resulta, de hecho, que toda la mate-



4. PRIMERA GENERACION de quarks y leptones. Forma un patrón ordenado cuando las partículas se acomodan de acuerdo con su carga eléctrica. Todos los valores de la carga entre +1 y −1 en intervalos de 1/3 quedan representados. Las partículas coloreadas poseen carga fraccionaria; las que carecen de color, carga entera. Este patrón es una característica arbitraria en el modelo standard, donde la carga y el color son independientes, pero podría tener alguna explicación si quarks y leptones fueran compuestos.



	PREON	CARGA ELECTRICA	COLOR	NUMERO DE GENERACION
GEUSTICONES	f_1	+1/2	SIN COLOR	0
	f_2	−1/2	SIN COLOR	0
CROMONES	C_R	+1/6	ROJO	0
	C_Y	+1/6	AMARILLO	0
	C_B	+1/6	AZUL	0
	C_0	−1/2	SIN COLOR	0
SOMONES	S_1	0	SIN COLOR	1
	S_2	0	SIN COLOR	2
	S_3	0	SIN COLOR	3

5. EN EL MODELO PREON se asignan tres propiedades de quarks y leptones a tres grupos de hipotéticos constituyentes, llamados geusticones, cromones y somones. Para formar un quark o un leptón se toma un preón de cada grupo. Los geusticones tienen responsabilidad principal en determinar la carga eléctrica, los cromones determinan el color y, los somones, el número de generación. En teoría, cada tipo de preón llevaría una propiedad, pero se precisan algunos ajustes para diferenciar la carga eléctrica fraccionaria de los quarks de la carga entera de los leptones. Aquí, los cromones portan carga eléctrica y color.

ria ordinaria puede formarse con un subconjunto que incluya sólo el quark *u*, el quark *d*, el electrón y el neutrino electrónico. Estas cuatro partículas y sus correspondientes antipartículas constituyen la "primera generación" de quarks y leptones; los restantes repiten el mismo patrón en otras dos generaciones, sin que parezcan añadir nada nuevo. Las partículas correspondientes en generaciones diferentes son idénticas en todos los aspectos, salvo en uno: tienen distinta masa. Los quarks *d, s* y *b*, por ejemplo, responden exactamente de la misma manera al electromagnetismo, al color y a la fuerza débil. Por alguna razón desconocida, sin embargo, el quark *s* pesa unas 20 veces más que el quark *d*, mientras el quark *b* es unas 600 veces más pesado que el *d*. Las fracciones entre las masas de los otros quarks y de los otros leptones cargados muestran un valor e injustificación parecidos (las masas de los neutrinos son demasiado pequeñas para haberse medido; no se sabe todavía si los neutrinos son muy ligeros o si carecen de masa por entero).

La presencia de tres generaciones de quarks y leptones demanda una explicación. ¿Por qué la naturaleza se repite a sí misma? También el patrón de las masas de las partículas está envuelto en el misterio. En el modelo standard, las masas quedan determinadas por unos 20 parámetros "libres", a los que puede asignarse cualquier valor que el teórico escoja; en la práctica, tales valores suelen fundarse en verificaciones experimentales. ¿No han de guardar los 20 parámetros ninguna relación? ¿Son, acaso, constantes fundamentales de la naturaleza, con el mismo status que la velocidad de la luz o la carga eléctrica del electrón? Probablemente no.

Otra desconcertante regularidad se aprecia en las cargas eléctricas de quarks y leptones: están relacionadas por fracciones simples y son todas ellas múltiplos enteros de la tercera parte de la carga del electrón. El modelo standard no da razón alguna de semejante fenómeno; en principio, los cocientes entre cargas podrían tener cualquier valor. De la observación se deduce que las fracciones un tercio y dos tercios que definen las cargas de los quarks no son meras aproximaciones. El protón consta de dos quarks *u* y uno *d*, con cargas 2/3 + 2/3 − 1/3, es decir, + 1. Si estos valores no fuesen exactos y, en cambio, los quarks tuviesen cargas, por ejemplo, + 0,617 y − 0,383, el valor de la carga del protón no sería exactamente igual a la del electrón, ni los átomos ordinarios serían eléctricamente neutros. Puesto que los átomos se encuentran juntos en número enorme, la más ligera desviación de su neutralidad se detectaría fácilmente.

Si ordenamos las partículas y antipartículas que forman una generación de acuerdo con su carga, hallaremos que cada valor desde −1 hasta +1, en intervalos de un tercio, está ocupado por una partícula (o, en el caso de carga cero, por dos partículas, precisamente el neutrino y el antineutrino). El patrón formado suscita todavía más preguntas. ¿Por qué la naturaleza ha privilegiado estos valores de la carga eléctrica y no otros, como +4/3 o −5/3? Es manifiesto que todas las partículas con carga entera carecen de color, mientras que las dotadas con carga fraccionaria lo tienen. ¿Media alguna relación entre la carga eléctrica de una partícula y su color, o entre quarks y leptones? El modelo convencional no implica tales relaciones, pero éstas parecen existir.

Otro motivo para buscar más allá del modelo convencional nace del permanente deseo de unificar las fuerzas fundamentales o, por lo menos, de encontrar alguna relación entre ellas. La concisión quedaría servida si, por ejemplo, se ensamblaran dos fuerzas, como lo fueron electricidad y magnetismo, o si una fuerza se convirtiera en residuo de otra, según aconteció con la fuerza fuerte con respecto a la fuerza de color. Pero, quién nos garantiza que una simplificación de este tipo no se alcance a costa de introducir nuevas fuerzas.

Una teoría que "trascienda" el modelo standard no debe contradecirlo ni invalidarlo. El modelo standard podría convertirse en un óptimo enfoque de la nueva teoría, más profunda. El modelo standard da una descripción bastante lograda de todos los fenómenos a distancias no inferiores a los 10^{-16} centímetros. Una teoría más profunda habría de abordar sucesos a distancias menores. Si quedan nuevos constituyentes por descubrir, existirán dentro de tan minúsculas regiones del espacio. Si hay nuevas fuerzas, su acción se dejará sentir sólo a distancias inferiores a 10^{-16} centímetros, sea la fuerza inherentemente de corto alcance (siguiendo el ejemplo de la fuerza débil) o esté sujeta a algún tipo de confinamiento (como lo está la fuerza de color).

La búsqueda de una teoría que trascienda el modelo convencional comenzó hace ya casi una decena de años; de entonces acá se han explorado varias direcciones. Una de ellas, muy prometedora, nos ha llevado hasta los modelos conocidos por teorías de gran unificación, que incorporan el electromagnetismo y las interacciones de color y débil en una fuerza fundamental. La idea básica consiste en poner todos los quarks y leptones que forman una generación en una sola familia; se postulan, entonces, nuevos bosones de aforo para mediar las interacciones entre quarks coloreados y leptones carentes de color. Estas teorías dan cuenta de las regularidades vistas en la distribución de la carga eléctrica y explican la conmensurabilidad de las cargas de los

quarks y leptones. Por otro lado, no hacen nada para reducir el número de constantes fundamentales, no arrojan luz sobre la triplicación de las generaciones y crean por sí mismas ciertas dificultades teóricas nuevas.

Ha habido diversas variaciones sobre el tema de la gran unificación. Así, el concepto de simetría horizontal aborda el problema de la triplicación estableciendo una relación de simetría entre las generaciones. La idea de la supersimetría, de gran belleza matemática, relaciona partículas con momento angular de espín semientero (quarks y leptones) con aquellas con espín entero (bosones de aforo). La teoría del technicolor sugiere que la partícula de Higgs del modelo convencional es un objeto compuesto formado por nuevas entidades fundamentales; estarían enlazadas por una nueva fuerza análoga al color, el technicolor. Estas ideas tratan de responder a otras tantas preguntas que permanecen abiertas en el modelo convencional. Pero cada idea fracasa también a la hora de contestar otras preguntas, trae nuevas dificultades y empeora algunas existentes; por ejemplo, incrementando el número de constantes arbitrarias no relacionadas.

En todos los esquemas para la gran unificación que acabamos de mencionar se da explícitamente por sentado que los quarks, los leptones, el fotón, los gluones y los bosones débiles son partículas fundamentales de la teoría última de la naturaleza. La alternativa de sugerir que los quarks y leptones son ellos mismos objetos compuestos es, en cierto sentido, la hipótesis más conservadora y menos original. Se trata de una estrategia que ha venido repitiendo su éxito: del átomo al núcleo, de éste al protón y, luego, al quark. Pero en otro sentido la idea de una subestructura para quarks y leptones es la propuesta más radical. El electrón se ha estudiado durante casi un siglo, y su naturaleza puntual está, de hecho, muy bien establecida. En el caso del neutrino, que quizás está totalmente desprovisto de masa, resulta más difícil todavía imaginar una estructura interna. Antes de afirmar con rotundidad que estas partículas y otras como ellas sean compuestas, habrá que superar obstáculos formidables.

Pero las dificultades de la empresa se saldarán con presumibles recompensas. Un modelo compuesto afortunado resolvería todas las preguntas que el modelo convencional deja pendientes. Tal hipotética teoría debería empezar introduciendo un nuevo conjunto de partículas elementales, a las que bautizaré prequarks. Lo ideal sería que no proliferaran. Cada quark y leptón del modelo standard se consideraría combinación de prequarks a semejanza de los hadrones, de los que se dice que son combinación de quarks. La masa de un quark o un leptón no se mostraría ya cual constante arbitraria de la naturaleza; antes bien, quedaría determinada por las masas de los prequarks constituyentes y la amplitud de la fuerza que une los prequarks. Las fracciones exactas que relacionan la carga de un quark con la de un leptón se explicarían de manera parecida: ambos tipos de partículas compuestas deducirían sus cargas a partir de las que poseyeran los mismos prequarks constituyentes. El patrón entero de quarks y leptones dentro de una generación reflejaría presumiblemente algunas reglas de combinación sencillas de los prequarks.

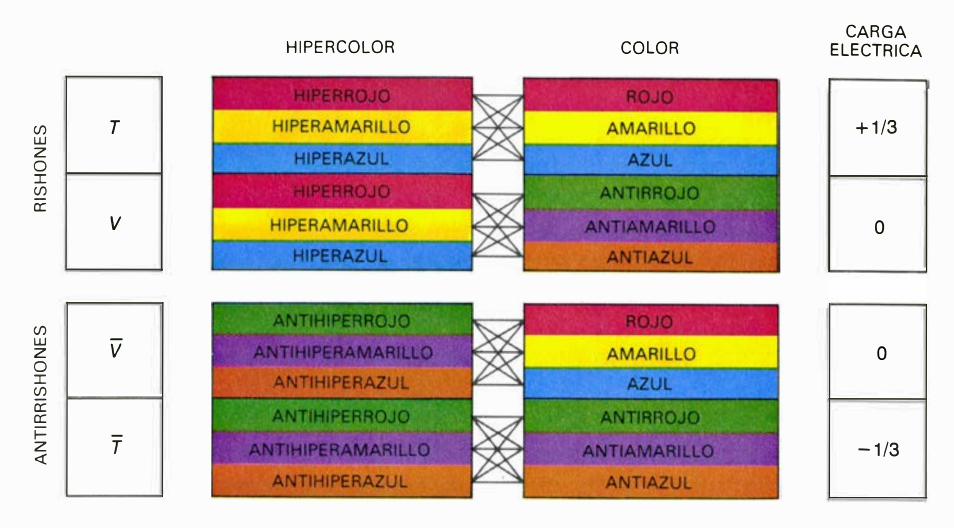
La existencia de múltiples generaciones recibiría también una explicación natural. Los quarks y leptones de una generación elevada tendrían una constitución interna similar a la correspondiente a las partículas de la primera generación; las diferencias estribarían en la energía y estado de movimiento de los constituyentes. Así, los quarks *s* y *b* serían estados excitados del quark *d*; el muon y el leptón tau, estados excitados del electrón. Se conocen ya estados excitados similares en los demás sistemas compuestos, incluidos átomos, núcleos y hadrones. Una docena de hadrones, por lo menos, se han identificado experimentalmente como estados excitados del protón. Se cree que tanto ellos como el propio protón se hallan integrados por una misma composición en quarks; a saber: *uud*.

Esta teoría prequark, imaginaria e ideal, cumple cuanto podríamos pedir, salvo la unificación de las fuerzas fundamentales. Cabría pensar algún progreso incluso aquí, puesto que ha-

6. COMBINACIONES DE PREONES dan lugar a los 24 quarks y leptones de las tres generaciones. Por ejemplo, el quark rojo *s* se construye a partir de un somón S_2 (señalando así que el compuesto constituye una partícula de la segunda generación), en combinación con el cromón rojo y el geusticón negativo.

brá que introducir, verosímilmente, una nueva fuerza para unir los prequarks; ésta quizá nos permitiera ahondar más en nuestro conocimiento sobre la relación que guardan las fuerzas hasta ahora aceptadas. Sin embargo, imaginar cómo debiera ser tan potente modelo no es lo mismo que construir uno real, dotado de coherencia interna. Por el momento nadie lo ha hecho.

Se echa en falta una teoría satisfactoria sobre la dinámica de los prequarks, una teoría que describiera cómo se mueven éstos dentro de un quark o un leptón y nos permitiera calcular la masa y energía total del sistema. Más adelante hablaremos de la existencia de obstáculos fundamentales, no insuperables es verdad, que se oponen a la formulación de tal teoría. Mientras tanto, y a falta de una descripción convincente del movimiento de los prequarks, los teóricos han explorado las posibilidades combinatorias de la idea prequark, es decir, han examinado los modos en que los quarks y leptones pueden construirse como combinaciones específicas de constituyentes más sutiles.



7. MODELO RISHON, desarrollado por el autor. Construye todos los quarks y leptones con sólo dos especies de partículas fundamentales y sus antipartículas. Los rishones llevan, amén de hipercolor, una propiedad asociada con la fuerza que los enlaza uno con otro, color ordinario, que transmiten a los sistemas compuestos que forman. Un rishón está cargado eléctricamente, en tanto que el otro es neutro.

A lo largo de los últimos años han ido apareciendo docenas de modelos. Quizá podríamos distribuirlos en cuatro o cinco grupos principales. Ningún modelo resuelve todos los problemas, satisface todas las preguntas, ni goza de la aceptación general. Tan injusto sería describir sólo un esquema cuan viable resultaría exponerlos todos. Repasaremos algunas ideas centrales.

El primer modelo explícito de una subestructura para quarks y leptones fue propuesto, en 1974, por Jogesh C. Pati, de la Universidad de Maryland en College Park, y Salam, que ha vuelto, desde entonces, al tema varias veces en colaboración con John Strathdee, del Centro Internacional de Física Teórica. A ellos se debe el término prequark, que he adoptado aquí como nombre genérico para todos los tipos de hipotéticos constituyentes. Las partículas elementales específicas del modelo ideado por Pati y Salam las llamaré preones, otro término de su cosecha.

El modelo preón arranca con la observación según la cual cada quark y leptón puede identificarse sin ambigüedades enumerando sólo tres propiedades: carga eléctrica, color y número de generación. Estas propiedades sugieren una manera directa de organizar un conjunto de partículas constituyentes. Se necesitan tres familias de preones. En una, los preones llevan carga eléctrica; en otra, color y, en la tercera familia, ostentan cierta propiedad que determina el número de generación. Para formar un quark o leptón, basta con seleccionar un preón de cada familia.

Los preones que determinan el número de generación se llaman somones, del griego soma, que significa cuerpo, en razón de su influencia dominante en la masa del sistema compuesto. Por haber tres generaciones de quarks y leptones, existirán tres somones. El color del sistema compuesto queda determinado por los preones llamados cromones: hay cuatro de ellos, uno de color rojo, otro amarillo, un tercero azul y el cuarto sin color. La familia de preones que queda, cuya misión consiste en definir la carga eléctrica, necesita tener sólo dos miembros para que cada quark y leptón resulte unívocamente identificado. Estos últimos preones reciben el nombre de geusticones (del griego γεύω, gustar saborear; debe preferirse a flavones, del inglés "flavor", de sinónimo significado); el caprichoso término designa la propiedad, sea la que sea, que distingue el quark *u* del *d*, el *c* del *s*, el neutrino del electrón, y así sucesivamente.

En el modelo preón, la clasificación de una partícula compuesta sígue-se directamente de su dotación de preones. Todos los leptones, por ejemplo, se distinguen por un cromón sin color, mientras todas las partículas de la primera generación tienen, obviamente, un somón de la primera generación. En la asignación de la carga eléctrica aparece, sin embargo, una complicación. Como hay sólo dos geusticones y son los únicos que llevan carga eléctrica, no todos los valores de la carga observados en la naturaleza se reproducirán. El quark *u* y el neutrino, por ejemplo, deben tener la misma carga (puesto que incluyen el mismo geusticón); lo mismo ocurre con el quark *d* y el electrón. El problema puede resolverse de varias maneras. En una de ellas, la carga eléctrica se asigna tanto a los geusticones como a los cromones; la carga total de la partícula compuesta es igual a la suma de los dos valores. Se pueden construir modelos de este tenor para conferir la carga correcta del estado considerado, a condición de que se abandone el principio según el cual cada tipo de preón transporta una sola propiedad.

Otro aspecto incómodo del modelo preón es el del requisito según el cual se formarán los compuestos sacando un único preón por familia. ¿Por qué no hay partículas formadas por tres cromones, por ejemplo, o por dos somones y un geusticón? Las exóticas propiedades de tales partículas las delatarían en seguida. Y parece verosímil que, caso de existir, se hubieran detectado ya.

Otros físicos han propuesto bastantes variaciones sobre el modelo preón, aceptando la misma idea fundamental aunque con conjuntos de preones ligeramente distintos. Entre dichas variaciones destacan los modelos sugeridos por Hidezumi Terazawa, Yoichi Chikashige y Keiichi Akama, de la Universidad de Tokio, y por O. Wallace Greenberg y Joseph Sucher, de la Universidad de Maryland.

En mi opinión, el modelo más sencillo de estructura de quarks y leptones

COMBINACION DE RISHONES	PARTICULA	COLOR	CARGA ELECTRICA
TTT	$\bar{e}$	SIN COLOR	+1
TTV	*u*	ROJO AMARILLO AZUL	+2/3
TVV	$\bar{d}$	ANTIRROJO ANTIAMARILLO ANTIAZUL	+1/3
VVV	ν_e	SIN COLOR	0
$\bar{V}\bar{V}\bar{V}$	$\bar{\nu}_e$	SIN COLOR	0
$\bar{V}\bar{V}\bar{T}$	*d*	ROJO AMARILLO AZUL	−1/3
$\bar{V}\bar{T}\bar{T}$	$\bar{u}$	ANTIRROJO ANTIAMARILLO ANTIAZUL	−2/3
$\bar{T}\bar{T}\bar{T}$	*e*	SIN COLOR	−1

8. COMBINACIONES DE RISHONES. Tomados de tres en tres (*columna de la izquierda*) dan razón cabal de todos los quarks y leptones (y antiquarks y antileptones) de cualquier generación. El patrón para las cargas eléctricas señalado en el modelo standard, y la relación aparente entre cargas fraccionarias y color, se suponen consecuencias naturales del modo en que los rishones se combinan. Todas las combinaciones permitidas de tres rishones o de tres antirrishones son neutras con respecto al hipercolor.

es el rishón, que presenté en 1979. Por aquellas fechas, Michael A. Shupe, de la Universidad de Illinois en Urbana-Champaign, lanzó una idea similar. Desde entonces, ha sufrido varios retoques y ha merecido el estudio detenido de Nathan Seiberg, y el mío propio, en el Instituto Weizmann de Ciencias en Rehovot. Se postulan sólo dos especies de bloques fundamentales constituyentes, llamados rishones. (*Rishon* es el adjetivo hebreo que significa primero o primario.) Un rishón tiene carga eléctrica +1/3, siendo el otro eléctricamente neutro. Los designaré, respectivamente, por *T* y *V*, iniciales de *Tohu Vavohu*, expresión hebrea cuya traducción castellana es "sin forma y vacío", descripción usada al comienzo del relato del Génesis para explicar el estado inicial del universo. Los antirrishones complementarios tienen cargas −1/3 y cero, y se designan por $\bar{T}$ y $\bar{V}$.

El modelo se ha dotado de una regla sencilla para construir quarks o leptones: tres rishones cualesquiera pueden ensamblarse y formar un sistema compuesto, o hacerlo tres antirrishones cualesquiera; ahora bien, no pueden mezclarse rishones y antirrishones en una sola partícula. Esa regla origina 16 combinaciones, que reproducen exactamente las propiedades de los 16 quarks, antiquarks, leptones y antileptones de la primera generación. En otras palabras, cada quark y leptón de la primera generación corresponde a una combinación permitida de rishones y antirrishones (en este sistema de clasificación cada color se cuenta por separado).

El patrón según el cual se generan las cargas de quarks y leptones es el siguiente. La combinación *TTT*, con cargas 1/3 + 1/3 + 1/3, lleva una carga total +1 y corresponde, por tanto, al positrón; similarmente $\bar{T}\bar{T}\bar{T}$ tienen carga total −1 y se identifica como el electrón. Las combinaciones *VVV* y $\bar{V}\bar{V}\bar{V}$, eléctricamente neutras ambas, representan al neutrino la primera y al antineutrino la segunda. Las restantes combinaciones permitidas dan lugar a los quarks dotados de carga fraccionaria. *TTV*, con una carga +2/3, es el quark *u*; *TVV*, con carga +1/3, el antiquark $\bar{d}$. Los estados análogos con antirrishones $\bar{V}\bar{V}\bar{T}$ y $\bar{V}\bar{T}\bar{T}$ corresponden al quark *d* y el antiquark $\bar{u}$.

El modelo da cuenta con igual éxito del color de los sistemas compuestos. Un rishón *T* puede adquirir cualquiera de los tres colores, rojo, amarillo o azul; un rishón *V*, un anticolor. Combinaciones del tipo *TTT* y *VVV*, que designan a los leptones, pueden convertirse en carentes de color, ya que admiten un rishón de cada color o uno de cada anticolor. Las otras combinaciones, que producen los quarks, tendrán un color neto. Así, un estado *TTV* mostrará los colores de los rishones rojo, azul y antiazul; el antiazul cancela el azul, dejando al sistema con un color rojo neto. De esta manera, la conexión entre color y carga eléctrica, manifiesta e injustificada en el modelo standard, recibe aquí una explicación satisfactoria. En virtud del modo en que se asigna la carga eléctrica y el color a los rishones, todos los sistemas compuestos con carga fraccionaria resultan coloreados y todos los sistemas con carga entera pueden tornarse "incoloros".

Otras regularidades del modelo standard pierden también su aura de misterio en cuanto se introducen los rishones. Consideremos el átomo de hidrógeno, constituido por un protón y un electrón; en términos de quarks y leptones: dos quarks *u*, un quark *d* y un electrón. El contenido total en rishones de dichos quarks son cuatro *T*, un $\bar{T}$, dos *V* y otros dos $\bar{V}$. La carga eléctrica del $\bar{T}$ cancela la carga de un rishón *T*, mientras que los *V* y los $\bar{V}$ se anulan también (en todo caso no tienen carga), dejando al protón con un contenido neto de carga igual al del sistema *TTT*. El contenido en rishones del electrón es el opuesto: $\bar{T}\bar{T}\bar{T}$. Se evidencia así la razón de que el protón y el electrón ostenten cargas de misma magnitud y de que el átomo de hidrógeno sea neutro: la fuente última de carga son pares igualados de partículas y antipartículas.

El modelo rishón, como otros muchos que explican el patrón de la primera generación, se halla en un aprieto a la hora de explicar la segunda y tercera generaciones. Parecería que tales modelos se prestasen bien al esquema según el cual cada partícula de las generaciones superiores se formaría como estado excitado de la correspondiente en la primera generación. Lo más sencillo sería describir el muon con los mismos prequarks constituyentes que el electrón, si bien en el muon los prequarks tendrían una configuración más energética. Aunque elegante, resulta desgraciadamente inviable. El esquema implica unas diferencias energé-

ticas, entre sucesivos estados excitados, mucho mayores que las diferencias reales. La deficiencia es fundamental y no parece tener remedio.

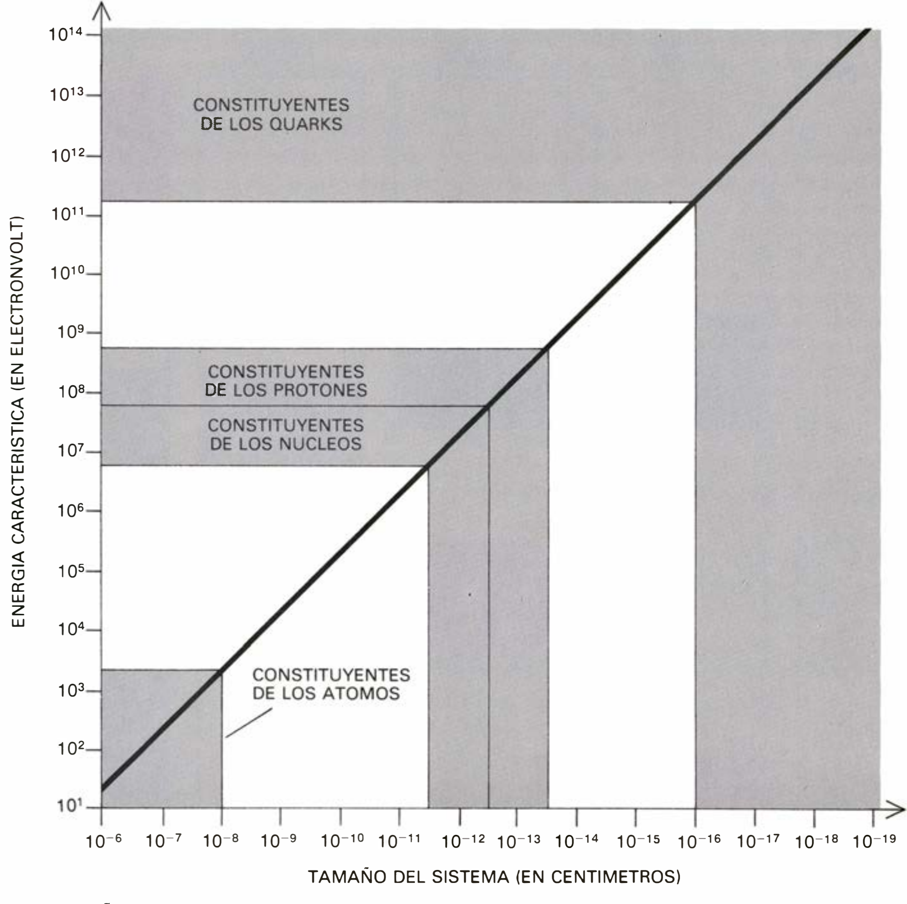
Se ha pensado en otro posible mecanismo para crear generaciones múltiples. Varios físicos han sugerido que las generaciones superiores correspondientes a un estado dado podrían crearse añadiendo una partícula de Higgs, la partícula "extra" asociada a los bosones débiles en el modelo standard. Puesto que la partícula de Higgs carece de carga eléctrica, color y momento angular de espín, añadida a un sistema compuesto alteraría sólo la masa. Un electrón podría convertirse entonces en un muon añadiendo una partícula de Higgs, o en una partícula tau agregando dos o más partículas Higgs. Seiberg y el autor avanzaron un mecanismo alternativo: una partícula de una generación superior podría formarse añadiendo pares de prequarks y antiprequarks. Todas las cargas y otras propiedades deben cancelarse en tal par, de manera que, de nuevo, sólo la masa quedaría afectada.

Nos movemos en un terreno meramente especulativo. Nadie sabe la razón de la distinción entre las tres generaciones, ni por qué son tres y no más. ¿O acaso son más? Nadie da explicación alguna de la diferencia de masas entre generaciones. En una palabra, la triplicación de las generaciones constituye, hoy por hoy, el principal rompecabezas con que se enfrentan los físicos.

Merece la pena que reparemos en un tercer tipo de modelo subestructural: aquel que pretende relacionar la posibilidad de una estructura para quarks y leptones con otro problema fundamental, a saber, el de la comprensión de la teoría cuántica relativista de la gravitación. De este tipo de cuestiones se han ocupado John Ellis, Mary K. Gaillard, Luciano Maiani y Bruno Zumino, del CERN. Para hacernos una idea de su enfoque, consideremos la distancia de interacción entre los prequarks: el límite experimental es inferior a los 10^{-16} centímetros, pero la distancia real podría ser inferior en varios órdenes de magnitud. Alrededor de los 10^{-34} centímetros, la fuerza gravitatoria adquiere intensidad suficiente para incidir de un modo significativo en las distintas partículas. Si la escala de la interacción entre prequarks alcanza ese reducido valor, no podrá despreciarse la gravitación. Ellis, Gaillard, Maiani y Zumino han trazado un ambicioso programa que pretende unificar todas las fuerzas, incluida la gravitación, en un esquema donde no sólo se abordan los quarks y leptones como partículas compuestas, sino que entran también en esa calificación los bosones de aforo. Mas, al igual que otros modelos compuestos, éste deja también muchos cabos por atar.

Detalles aparte, todo modelo prequark debe aportar algún mecanismo que los enlace. Ha de mediar una potente fuerza atractiva entre ellos. Se puede seguir la estrategia de postular una nueva interacción fundamental de la naturaleza, análoga en sus resultados a la fuerza de color del modelo convencional. Para destacar semejante analogía, se ha puesto a la nueva fuerza el nombre de hipercolor, y de hipergluones los campos portadores. Se supone que los prequarks tienen hipercolor, aunque se combinan para formar sistemas compuestos carentes de hipercolor, de la misma manera que los quarks tienen color ordinario pero se combinan formando protones y neutrones sin color. Presumiblemente, la fuerza de hipercolor da lugar a la propiedad del confinamiento, otra vez en analogía con la fuerza de color. Así, todos los prequarks hipercoloreados estarían encerrados dentro de partículas compuestas, lo que explicaría por qué no se ven prequarks libres en los experimentos. De una idea de este tipo hablaba 't Hooft, quien estudió algunas de sus consecuencias matemáticas y no se abstuvo de expresar sus dudas de que la naturaleza siguiera ese camino.

El radio típico del confinamiento del hipercolor debe ser inferior a 10^{-16} centímetros. Sólo sondeando la materia a distancias inferiores a ésta cabría ver los hipotéticos prequarks y sus hipercolores. A un rango de 10^{-14} o 10^{-15} centímetros, el hipercolor desaparece; los únicos objetos visibles a tal escala de resolución (los quarks y leptones) son neutros respecto al hipercolor. A un rango de 10^{-13} centímetros, el co-



9. TAMAÑO Y ENERGIA tienen una relación recíproca en la teoría cuántica, lo que significa que los constituyentes de los quarks y leptones compuestos deben tener una energía cinética sumamente elevada. El tamaño de un átomo implica que sus constituyentes porten energías desde unos cuantos electronvolt hasta millares de ellos (un electronvolt es la energía que gana un electrón acelerado bajo una diferencia de potencial de un volt). En el interior de un núcleo los protones y neutrones se mueven con una energía de varios millones de electronvolt; en un protón o un neutrón, los quarks tienen energías de varios centenares de millones de electronvolt. Cualquier constituyente de quarks y leptones debe estar confinado en un radio de menos de 10^{-16} centímetros; posiblemente, mucho menor. Resulta así que la energía cinética de los hipotéticos prequarks no será inferior a varios centenares de miles de millones de electronvolt.

lor ordinario se extingue a su vez, y el mundo se nos ofrece formado por objetos incoloros e hiperincoloros: protones, neutrones, electrones y demás.

La noción de hipercolor encaja bien en un amplio despliegue de modelos prequark, incluido el modelo rishón. Amén de su carga eléctrica y su color, se supone que los rishones poseen hipercolor y, los antirrishones, antihipercolor. Sólo están permitidas las combinaciones de tres rishones o tres antirrishones, porque sólo tales combinaciones son neutras respecto al hipercolor. No cabe ningún sistema mixto de tres partículas, $TT\bar{T}$ por ejemplo, puesto que no carecería de hipercolor. La asignación de hipercolores fundamenta, pues, la regla para formar sistemas compuestos de rishones. En los otros modelos prequark basados en el hipercolor se aplican reglas similares.

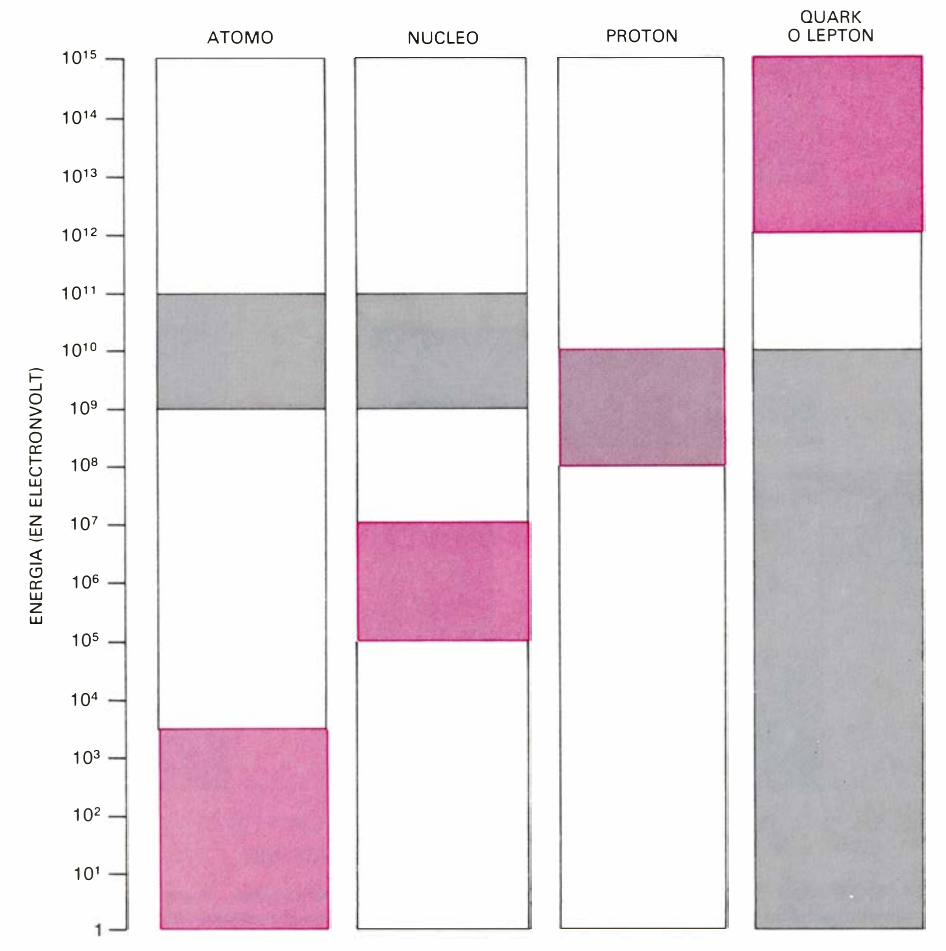
El modelo prequark pretende simplificar nuestra comprensión de la naturaleza, de ahí que postular una nueva fuerza básica no parece servir de mucho. Pero en el caso del hipercolor quizás encontremos alguna compensación. Consideremos el neutrino: sin carga eléctrica ni color; sólo posee carga débil. De acuerdo con el modelo standard, dos neutrinos actuarán uno en el otro a través de la fuerza débil de corto alcance. Si los neutrinos constaran de prequarks hipercoloreados, habría una fuente adicional de interacciones entre ellos. Cuando dos neutrinos se hallaran muy distantes, no mediaría prácticamente fuerza de hipercolor entre los mismos; pero cuando se acercaran mucho, los prequarks hipercoloreados dentro de un neutrino podrían "ver" los hipercolores internos del otro. Emergerían entonces complicadas atracciones y repulsiones de corto alcance. Nos encontramos ante el mismo mecanismo que explica la fuerza molecular como residuo de la electromagnética y la fuerza fuerte como residuo de la fuerza de color.

La conclusión puede ser también la misma. Seiberg y el autor por un lado, y Greenberg y Sucher por otro, fueron los primeros en sugerir que la fuerza débil de corto alcance podría constituir un efecto residual de la fuerza de hipercolor. De acuerdo con dicha hipótesis, los bosones débiles W^+, W^- y Z^0 deben ser, también, objetos compuestos, formados, presumiblemente, por combinación de los mismos prequarks que constituyen quarks y leptones. Si esta idea se confirma, la lista de fuerzas fundamentales tendrá cuatro apartados: gravitación, electromagnetismo, color e hipercolor. Debe notarse, sin embargo, que todas estas fuerzas son de largo alcance; las de corto alcance (molecular, fuerte y débil) habrán perdido su carácter fundamental.

Por ahora, el hipercolor se mueve en el plano de la conjetura; lo mismo ocurre con la explicación según la cual la fuerza débil es residuo del hipercolor. Todavía podría caber que la fuerza débil fuese fundamental. Medidas cuidadosas de la masa, vida media y otras propiedades de los bosones débiles arrojarán luz sobre el tema.

El hipercolor no es el único candidato a erigirse en fuerza de enlace de los prequarks. De otra posibilidad interesante hablaron Pati, Salam y Strathdee. En vez de introducir una nueva fuerza de hipercolor, recogieron una idea añeja en física, la de la fuerza magnética, y la adaptaron al nuevo propósito. Un imán ordinario tiene, invariablemente, dos polos, que pueden considerarse cargas magnéticas opuestas. A lo largo de cincuenta años se han ido aduciendo razones teóricas para abonar una posible existencia de cargas magnéticas aisladas, los monopolos. Pati, Salam y Strathdee han sostenido que los prequarks podrían ser partículas con cargas parecidas a las magnéticas y a las eléctricas. Si es así, las fuerzas que los unen pueden tener un origen nuevo.

Ninguna de las ideas que acabo de exponer constituye una teoría de la dinámica de los prequarks. La verdad es que nos hallamos con un grave obstáculo que impide formular tal teoría; se requieren unos prequarks sutilísimos. El límite más riguroso para su tamaño se obtiene indirectamente con las medidas del momento magnético del electrón, que concuerda con los cálculos de la electrodinámica con una precisión de diez cifras significativas. En el cálculo se supone que el electrón es puntual; si tuviese cierta extensión espacial o estructura interna, el valor medido diferiría del calculado. Evidentemente, tal discrepancia ha de afectar, como máximo, al onceavo dígito del resultado. Esta restricción implica que la

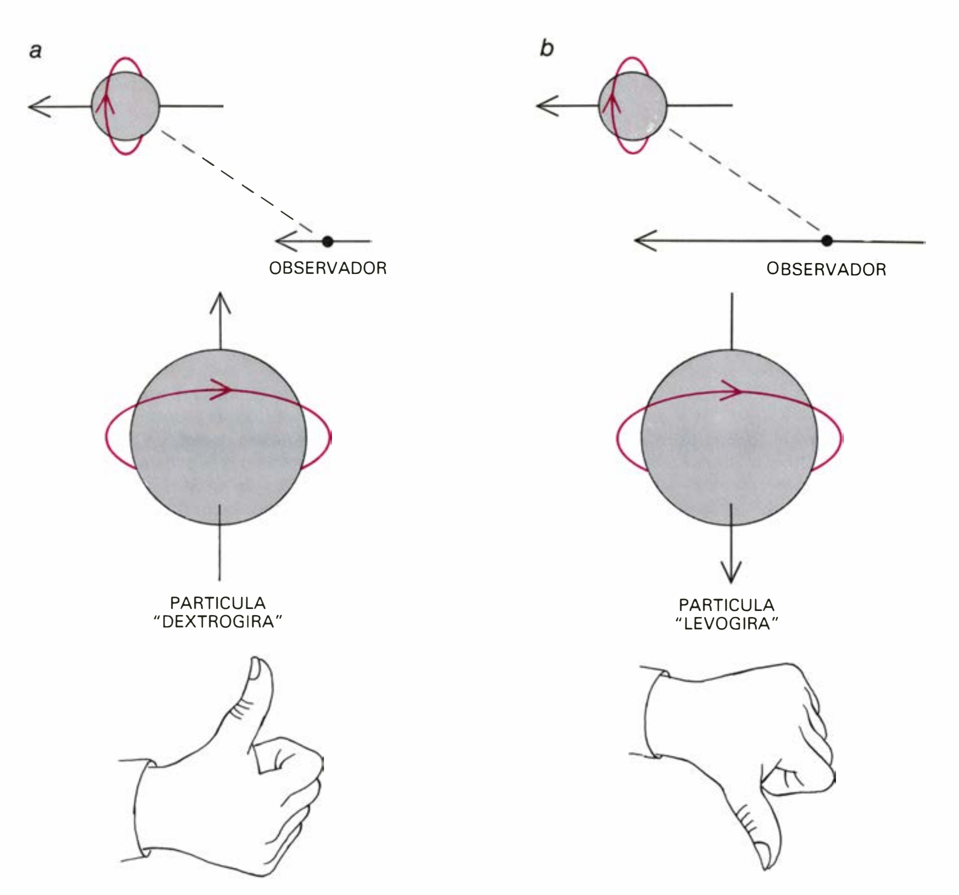


10. ANTE LA DISCORDANCIA ENTRE ENERGIA Y MASA se hace difícil imaginar una teoría sobre el comportamiento e interacción entre prequarks. En un átomo, o en un núcleo, la energía cinética de sus constituyentes (*color*) es mucho menor que la masa total del sistema (*gris*). En un protón, las dos cantidades poseen una magnitud comparable. En un quark compuesto, sin embargo, la energía de los prequarks excede en mucho la masa total. Comparada con la energía cinética, la masa es, en esencia, nula. De un modo u otro, todas las masas se cancelan virtualmente, lo que no es verosímil que resulte accidental.

escala característica de distancias en la estructura interna del electrón sea inferior a 10^{-16} centímetros. En lenguaje llano, éste es el radio máximo de un electrón, y cualquier prequark deberá hallarse en su interior. Si gozaran de una dispersión mayor, su presencia se hubiera detectado ya.

¿Por qué el exiguo tamaño del electrón prohíbe especular sobre su estructura interna? El principio de incertidumbre establece una relación recíproca entre el tamaño de un sistema compuesto y la energía cinética de cualquier componente que se mueva por su interior. Cuanto menor sea el sistema, mayor será la energía cinética de los constituyentes. Se sigue que los prequarks habrán de tener enormes energías: mayores que 100 GeV (100.000 millones de electronvolt), y, posiblemente, mucho más (un electronvolt es la energía adquirida por un electrón cuando se acelera bajo una diferencia de potencial de un volt). Como la masa es fundamentalmente equivalente a la energía, puede medirse en el mismo sistema de unidades. Así, por ejemplo, la masa del electrón es equivalente a 0,0005 GeV. Hay aquí una paradoja, que llamaré discordancia energética: la masa del sistema compuesto (si lo es realmente) es mucho menor que la energía de sus constituyentes.

Para ilustrar la singularidad de esta situación, pensemos en las relaciones entre masa y energía cinética en otros sistemas compuestos. En un átomo, la energía cinética de un electrón es inferior a la masa del átomo en muchos órdenes de magnitud. En el hidrógeno, por ejemplo, la razón es, aproximadamente, de una parte en cien millones. La energía que se necesita para cambiar la órbita del electrón y poner así al átomo en un estado excitado es también una fracción despreciable de la masa atómica. En un núcleo, la energía cinética de protones y neutrones es también pequeña, en comparación con la masa nuclear, aunque no hasta el punto de despreciarla. El movimiento de las partículas les confiere una energía cercana al uno por ciento de la masa del sistema. La energía necesaria para crear un estado excitado es también del uno por ciento de la masa, aproximadamente.

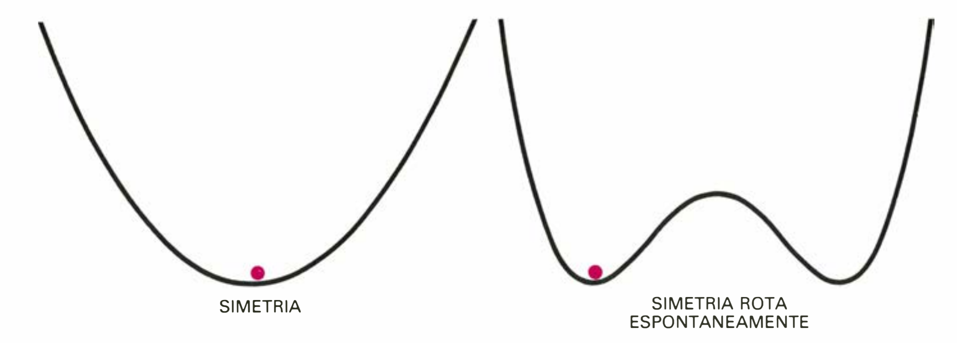


11. LA SIMETRIA QUIRAL ofrece una posible explicación para la "milagrosa" cancelación de la masa en quarks y leptones. La quiralidad, o helicidad, describe la relación del momento angular de espín de una partícula con su dirección de movimiento. Supongamos que al observador lo alcanza un electrón que se mueve más rápido (*a*). Según el punto de vista del observador, el electrón obedece la regla de la mano derecha: cuando los dedos de esa mano giran en la misma dirección del espín, el pulgar da la dirección del movimiento. Sin embargo, si el observador aumenta su velocidad, para alcanzar al electrón (*b*), la helicidad de la partícula cambia. En el sistema de referencia del observador el electrón ahora se acerca en vez de alejarse, pero la dirección de su espín no ha cambiado; de ello resulta que su movimiento queda descrito por una regla de mano izquierda. En consecuencia, no se conserva la quiralidad. Hay un tipo de partículas para las que no puede aplicarse este razonamiento, a saber, partículas sin masa, que deben moverse siempre con la velocidad de la luz. Ningún observador se moverá más rápido que una partícula sin masa; su helicidad es una propiedad invariante. Si una teoría de prequarks presenta simetría quiral, en la cual la helicidad debe conservarse, la pequeña masa de quarks y leptones quizá no sea accidental. Si se quiere que se mantenga la simetría quiral, tendrán que carecer, casi por entero, de masa.

Con el protón y sus quarks constituyentes la relación energía-masa empieza a adquirir tintes curiosos. A partir del radio eficaz del protón puede calcularse la energía típica de sus quarks componentes; resulta comparable a la propia masa del protón, que es un poco inferior a 1 GeV. La energía que debe gastarse para crear un estado excitado de un sistema de quarks es del mismo orden de magnitud: la masa de los hadrones identificados como estados excitados del protón superan la de éste en un 30 o 100 por ciento. Aún así, el cociente de la energía cinética respecto a la masa total está todavía en el rango que parece intuitivamente razonable. Supongamos que se conociera sólo el radio del protón y, consiguientemente, la energía típica de lo que hubiera encerrado en él, y quisiéramos averiguar la masa del protón. Puesto que la energía de los constituyentes es, generalmente, de unos pocos centenares de millones de electronvolt, pensaríamos que la masa total del sistema es, por lo menos, del mismo orden de magnitud o quizá mayor. Sería una suposición correcta.

Para el átomo, el núcleo y el protón, la masa del sistema es, como mínimo, de un valor igual al de la energía cinética de los constituyentes; en algunos casos, mayor. Si los quarks y leptones están compuestos, la relación de la energía respecto de la masa debe ser, sin embargo, bastante diferente. Como los prequarks tienen energías por encima de los 100 GeV, se deduciría que forman compuestos con masas de centenares de GeV o más. Los quarks y leptones conocidos tienen una masa mucho menor; en el caso del electrón y los neutrinos la masa es inferior, por lo menos, en seis órdenes de magnitud. El total es inferior, de lejos, a la suma de sus partes.

La alta energía de los prequarks excluye también la idea de mirar las generaciones superiores de quarks y leptones como estados excitados de los mismos conjuntos de prequarks que



12. RUPTURA ESPONTANEA DE SIMETRIA, un mecanismo que puede desbaratar una teoría prequark, aun cuando goce de simetría quiral. Los sistemas aquí dibujados –un simple cuenco y un cuenco con un tope en el fondo– pueden describirse como simétricos: intercambiando derecha e izquierda el sistema permanece inalterado. Para el simple cuenco, el sistema es aún simétrico si le ponemos una bola; ésta permanece en reposo en el centro, de modo que cambiar derecha e izquierda no tiene ningún efecto. En el cuenco con el tope, la bola toma asiento en un lado u otro, y la simetría queda inevitablemente rota. Similarmente, una teoría prequark que tenga simetría quiral puede producir sistemas compuestos que que no siguen la simetría. Mostrar que una teoría quiral puede quedar sin ruptura constituye hoy el principal desafío al que se enfrenta la formulación de una teoría sobre el movimiento de los prequarks.

forman las partículas de la primera generación. Al igual que en los demás sistemas compuestos, la energía necesaria para cambiar las órbitas de los prequarks debería ser del mismo orden de magnitud que la energía cinética de sus constituyentes. Sería, pues, de esperar que las sucesivas generaciones difiriesen en masa por centenares de GeV, mientras que las diferencias reales de masa son mínimas, de 0,1 GeV.

En este momento, podríamos admitir la inviabilidad de la discordancia energética; de hecho, la discordancia demostraría simplemente la elementaridad y la naturaleza carente de estructura de los quarks y leptones. Muchos físicos sostienen este punto de vista. Sin embargo, la discordancia energética no contradice ninguna ley básica de la física; y nada me impide defender que la prueba circunstancial del carácter compuesto de quarks y leptones es suficientemente persuasiva para garantizar investigaciones adicionales.

Lo peculiar de la masa de quarks y leptones no es tanto su sutileza cuanto el hecho de que sea virtualmente cero cuando la medimos en la escala energética definida por la energía de sus constituyentes. En otros sistemas compuestos, una pequeña cantidad de masa se "pierde" convirtiéndose en energía de enlace del sistema. La masa total de un átomo de hidrógeno, por ejemplo, es ligeramente menor que la de un protón y electrón aislados; la diferencia equivale a la energía de enlace. En un núcleo, este "defecto de masa" alcanza un escaso porcentaje de la masa total. En un quark o leptón la masa entera del sistema se cancela, al parecer, casi exactamente. Aunque una cancelación tan "milagrosa" no sea imposible, resulta improbable que ocurra por azar. Se conocen grandes cancelaciones similares en otras partes de la física, siempre derivaron de algún principio de simetría o de alguna ley de conservación. Para abrigar esperanzas de construir una teoría sobre dinámica de prequarks se impone encontrar tal simetría en este caso.

Existe un candidato probable: la simetría quiral, o quiralidad. La palabra deriva del nombre griego de mano, y la simetría tiene que ver con la helicidad, la propiedad definida por el espín de la partícula y la dirección del movimiento. Como otras simetrías de la naturaleza, la quiral lleva asociada una ley de conservación, ley que da la explicación más clara sobre el significado de la simetría. La ley establece que el número total de partículas dextrógiras y el número total de levógiras no pueden cambiar nunca.

En el mundo ordinario de los protones, electrones y partículas similares, la helicidad o quiralidad no se conserva. Un sencillo experimento imaginario va a mostrarnos un caso de violación de esta ley de conservación. Pensemos en un observador que se mueva en una línea recta, mientras lo alcanza un electrón. Al alejarse el electrón, advierte él que su espín y dirección de movimiento están relacionados por una regla de mano derecha. Supongamos ahora que el observador aumenta su velocidad para alcanzar al electrón. En la referencia del observador el electrón parece acercarse; en otras palabras, tiene sentido opuesto al de antes. Ahora bien, su espín no ha cambiado; se ha convertido, pues, en una partícula levógira.

Hay un tipo de partículas que no admiten este experimento imaginario: las carentes de masa. Estas deben moverse siempre a la velocidad de la luz; ningún observador adelantará, pues, su paso. De ello resultará que la helicidad de una partícula sin masa será una propiedad invariante, independiente del sistema de referencia del observador. Puede demostrarse, además, que ninguna de las fuerzas de la naturaleza conocidas (las mediadas por fotones, gluones y bosones débiles) altera la helicidad de una partícula. Si el mundo estuviera hecho exclusivamente de partículas carentes de masa, podría decirse que el mundo tendría simetría quiral.

La simetría quiral es la raíz de una idea que podría dar cuenta de la pequeña masa de quarks y leptones. Por la razón siguiente. Si los prequarks son partículas sin masa, poseen espín 1/2 e interactúan entre sí sólo a través del intercambio de bosones de aforo, queda garantizado que cualquier teoría que describa su movimiento tendrá una simetría quiral. Si los prequarks sin masa se enlazan para formar objetos compuestos de espín 1/2 (a saber, los quarks y leptones), la simetría quiral puede asegurar que las partículas compuestas poseerán también masa nula comparada con la enorme energía de los prequarks encerrados en ellas. Por tanto, la exigua masa de quarks y leptones no es un accidente. Carecerán de ella con respecto a las energías de sus constituyentes, si queremos que la simetría quiral de la teoría se mantenga.

El punto crucial de esta argumentación es el que extiende la simetría quiral de un mundo de prequarks carentes de masa a otro integrado por quarks y leptones compuestos. Es imprescindible que la simetría del sistema físico original sobreviva y la respeten los sistemas compuestos formados con los constituyentes sin masa. Puede parecer evidente que si una teoría es, en cierto sentido, simétrica, los sistemas físicos que describe exhiban dicha simetría; la verdad es que la ruptura espontánea de simetrías resulta un fenómeno corriente. La ruleta nos ofrece un ejemplo familiar. Una teoría física de la ruleta mostraría que es completamente simétrica en el sentido de que cada lugar es equivalente a cualquier otro. Pero el sistema físico que resulta de introducir una bola en la ruleta es asimétrico: la bola, invariablemente, queda en reposo en un solo lugar.

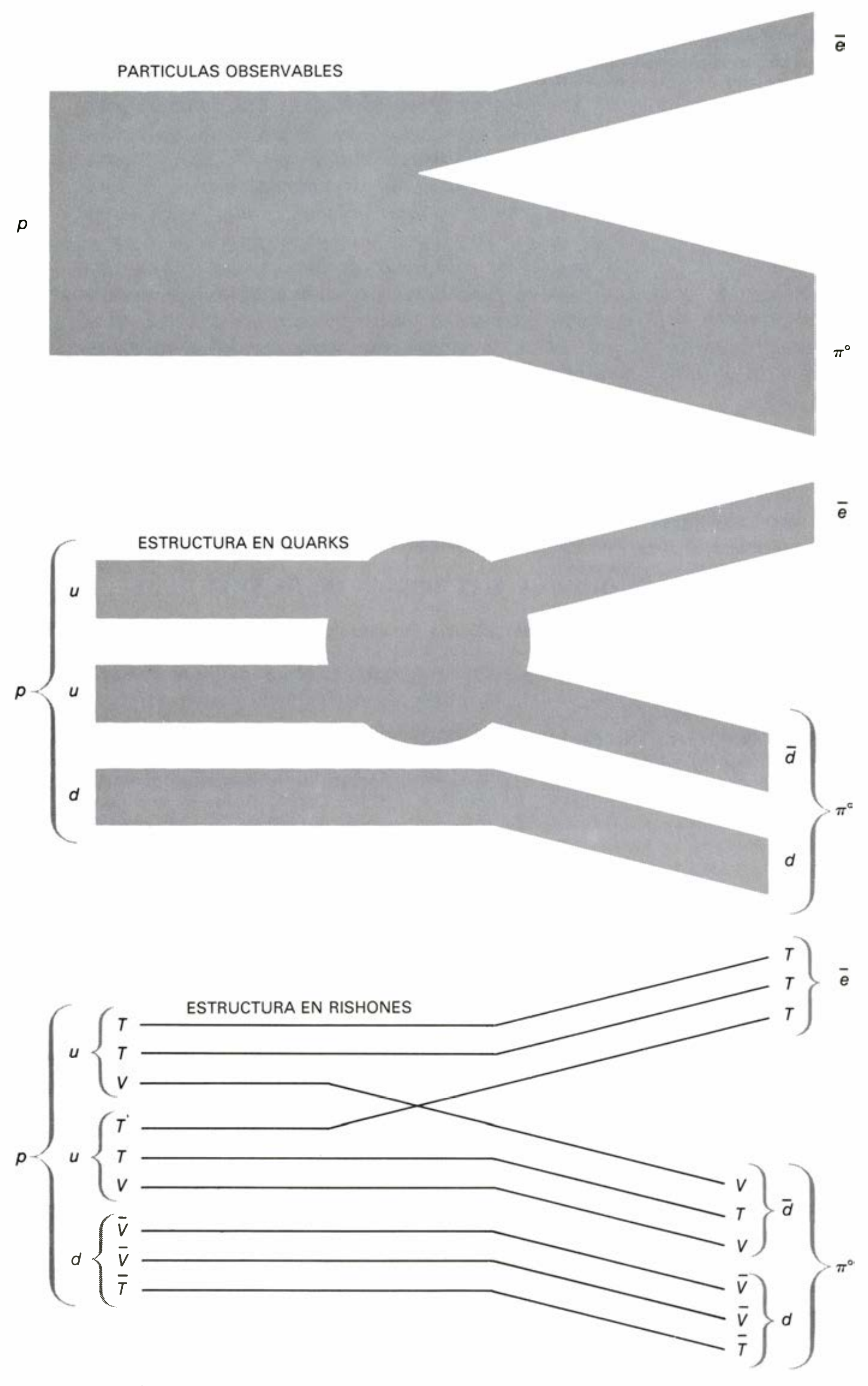
En el modelo standard la ruptura espontánea de una simetría es la que hace masivos a los tres bosones débiles y deja sin masa al fotón. La teoría que

describe estos bosones de aforo es simétrica; en ella, los cuatro bosones son esencialmente indistinguibles, mas en virtud de la ruptura de simetría los estados físicos observados difieren por entero. Las simetrías quirales son notoriamente susceptibles a la ruptura de simetría. Si la simetría quiral de los prequarks llega o no a romperse cuando los prequarks forman objetos compuestos, sólo puede determinarse a través de una comprensión detallada de las fuerzas que actúan sobre los prequarks. Por ahora, tal comprensión no existe. En ciertos modelos se aprecia la existencia de una tal simetría quiral, pero está claramente rota. Nadie ha tenido éxito aún construyendo un modelo compuesto de quarks y leptones en el que se sepa que la simetría quiral permanece sin romper. Ni el modelo preón ni el modelo rishón logran resolver el problema. Esta tarea es, probablemente, la más difícil de cuantas encuentran quienes se aprestan a demostrar que quarks y leptones son objetos compuestos.

Si se construye una teoría prequark coherente, tendrá que pasar aún la comprobación experimental. Importa, en primer lugar, establecer en el laboratorio si los quarks y leptones tienen o no estructura interna. En caso afirmativo, los experimentos empezarán a seleccionar entre los diversos modelos. Los experimentos habrán de penetrar en un reino desconocido de distancias inferiores a los 10^{-16} centímetros y energías superiores a los 100 GeV. Hay dos maneras básicas de explorar esta región: ensayando con partículas aceleradas a muy altas energías o midiendo, con precisión, cantidades a baja energía que dependan de la física de los sucesos a muy cortas distancias.

Entre los experimentos del primer tipo se inscriben la investigación sobre los bosones débiles y la búsqueda de las partículas de Higgs del modelo standard. Cuando puedan producirse tales partículas en número suficiente, una ojeada cuidadosa a sus propiedades revelará mucho acerca de la física a muy cortas distancias. Se espera que los nuevos aceleradores en fase de proyecto o construcción en los Estados Unidos, Europa y Japón produzcan información detallada sobre los bosones débiles y continúen también los estudios iniciados sobre los propios quarks y leptones.

Parejo interés encierran los experimentos de baja energía con alta precisión. Uno de ellos es la búsqueda de la desintegración del protón, partícula cuya vida media dura, por lo menos, 10^{30} años. Hay varios despliegues experimentales que están sometiendo a control cantidades ingentes de materia, incorporando más de 10^{30} protones, en un esfuerzo por detectar señales emitidas cuando se desintegra uno de éstos. Ninguna de las fuerzas del modelo standard puede inducir tal suceso, pero ninguna de las reglas de dicho modelo lo prohíbe en absoluto. Por otro lado, tanto las teorías de gran unificación como los modelos prequark comprenden



13. DESINTEGRACION DEL PROTON: un suceso supuesto que puede interpretarse como prueba experimental para una teoría de gran unificación o para un modelo de subestructura de los quarks. En un tipo de desintegración, el protón se desdobla en un positrón ($\bar{e}$) y un pion neutro (π^0). El suceso puede entenderse en función de los quarks constituyentes del protón: una interacción entre los dos quarks *u* convierte a uno de ellos en un positrón y al otro en un antiquark $\bar{d}$; este último se combina con el *d* restante del protón para formar el pion neutro. Las teorías de gran unificación sugieren que la interacción de los quarks *u* está mediada por una nueva fuerza de la naturaleza. El modelo rishón, diseñado por el autor, proporciona una explicación alternativa, a saber: los dos quarks *u* intercambian un rishón *T* y uno *V*.

mecanismos que convierten un protón en otras partículas que, por último, dejarían tras de sí sólo leptones y fotones. Si la desintegración se detecta, su proporción y el patrón de los productos de desintegración ofrecerían una importante perspectiva más allá del modelo standard.

Hay un interés similar en los procesos hipotéticos en los que un muon emite un fotón y se convierte entonces en un electrón. De nuevo, ninguna fuerza del modelo standard puede ocasionar un suceso de este tipo; de nuevo, ninguna ley fundamental lo prohíbe tampoco. Algunos de los modelos compuestos permiten esta transición; no así otros. Por consiguiente, la búsqueda de este proceso puede ofrecer un medio para escoger entre los modelos. Los experimentos realizados hasta el presente ponen un límite por debajo de uno entre 10.000 millones para la probabilidad de que un muon dado se desintegre de esta suerte. La detección de tal suceso y la determinación de su proporción arrojaría luz sobre la misteriosa distinción entre generaciones.

Hay un tercer tipo de experimentos de precisión: los que continúan refinando las medidas del momento magnético del electrón y del muon. Pueden esperarse mejoras adicionales tanto en la precisión experimental cuanto en los correspondientes cálculos en electrodinámica cuántica. Si los resultados continúan en concordancia con las predicciones del modelo standard, el límite para el posible tamaño de la subestructura de quarks y leptones se alejaría todavía más. Si se detectara una discrepancia entre teoría y experimento, habríamos obtenido una prueba sólida de que los quarks y leptones no son elementales.

Quizás hayan de transcurrir diez o veinte años antes de que el siguiente nivel, en la estructura de la materia, se nos aparezca (en la hipótesis de que existiera otro nivel). Lo que se necesita es una descripción teórica firme, que sea coherente en sí misma, que concuerde con todos los experimentos y de sencillez suficiente para explicar todos los rasgos del modelo standard a través de unos cuantos principios, pocos, y de unas cuantas partículas y fuerzas fundamentales, muy pocas también. La descripción correcta, se trate de una teoría de gran unificación o se trate de un modelo compuesto de quarks y leptones, quizás exista ya en forma embrionaria. ¿Por qué no admitir que la teoría saldrá de una idea totalmente nueva? Como diría Niels Bohr, tal vez nuestras actuales ideas "no sean suficientemente disparatadas para ser correctas".

Ciencia y sociedad

Un tema envenenado

Han pasado dos años desde que surgió en España el brote epidémico designado como "síndrome tóxico". Esta efemérides parece justificar algunas reflexiones sobre lo sucedido, teniendo en cuenta especialmente, por lo que a nosotros respecta, sus aspectos científicos.

A pesar de lo modesto del empeño, llevarlo a cabo ha resultado muy arduo. Hasta ahora no se dispone de ninguna explicación global del asunto hecha por órgano competente, si se exceptúa el Informe elaborado por el Gobierno con motivo del debate parlamentario de septiembre de 1981, cuando, si bien la enfermedad había perdido ya virulencia desde el punto de vista puramente sanitario, la polémica y el escándalo público que desde sus inicios le han rodeado estaban en su cenit. Una vez producida la insólita situación sanitaria, administrativa y científica de que un Estado se hiciera prácticamente responsable de una epidemia, parecería lógico que al menos se hubiese continuado informando periódicamente de su evolución con documentos similares a los ofrecidos en septiembre del 81. Pero esto no ha sucedido, a pesar de haberse realizado entre tanto una renovación completa de los poderes legislativo y ejecutivo, de haber proliferado las comisiones, interministeriales y de muchos otros tipos, los grupos de seguimiento, e incluso de haberse creado órganos específicos de la administración dedicados exclusivamente al mismo. Quien quiera comprobar la pobreza y vaguedad de lo que se sabe oficialmente al respecto haría bien en leer, como ejemplo, el informe elaborado por la ponencia del Senado correspondiente a la Comisión Especial de Investigación del Síndrome Tóxico con fecha 21 de abril de 1983, después de recabar, suponemos, cuanta información considerase necesaria. El resumen de lo deducido de las comparecencias realizadas y de los informes presentados ante la Comisión sobre los aspectos clínicos y de investigación ocupa apenas un folio mecanografiado (más dos y medio de anexos).

Se dispone, sin embargo, de información de otras fuentes y en este caso hay que confesar que abundante y detallada. No menos de dos mil páginas ocupan los informes publicados sobre el particular por el Instituto de la Grasa de Sevilla, por el Ministerio de Sanidad y Consumo con motivo del Simposium nacional de junio de 1982 y por el Consejo Superior de Investigaciones Científicas como resultado de su programa para el estudio del síndrome tóxico en marzo de 1983. Si excluímos las nueve páginas escritas por el Coordinador de este último programa, Angel Pestaña, como introducción al primero de los dos volúmenes del CSIC, las demás contienen diversos trabajos de investigación sobre aspectos muy circunscritos de la enfermedad, ya sean de tipo médico (clínico, patológico, epidemiológico, etc.), analítico o bioquímico, pero ningún intento de síntesis explicativa utilizable. Hay millones de datos, abundancia de hipótesis, numerosas cautelas, vagas quejas sobre deficiencias e incertidumbres, muchas afirmaciones superespecializadas sobre aspectos muy concretos del problema, pero quien quiera entender el porqué y el cómo de lo que sucedió tendrá que seleccionar y evaluar los datos que considere pertinentes, organizarlos en un argumento y sacar sus propias conclusiones. Tropezará para ello con las siguientes dificultades, por lo menos.

En primer lugar, la falta de datos. Esto no es incompatible con lo que acabamos de decir. Los datos contenidos en las publicaciones indicadas son de carácter muy parcial y no mencionan siquiera áreas enteras del problema. Baste un ejemplo. Durante el mes de mayo de 1981, al iniciarse la epidemia, se atribuyó su causa a una infección por bacterias o por virus. Los trabajos realizados para aislar el agente obtuvieron, según se dice en algunas partes, resultados más bien negativos, pero no sabemos que hayan sido publicados, ya que desde el día 10 de junio se aceptó de modo unánime y exclusivo la hipótesis de que la enfermedad era de carácter tóxico. Sobre esta hipótesis se concentraron masivamente desde entonces todos los esfuerzos y, con mayor o menor convicción, todos los trabajos publicados la toman por punto de partida. Por muy fundada que estuviera la hipótesis tóxica, ello no debería haber excluido una discusión más profunda de la infecciosa, entre otras razones porque cuanto más clara fuese la demostración de que no era válida más justificado estaría insistir en otras. En los momentos actuales, cuando la hipótesis tóxica tampoco tiene el grado de confirmación que cabría esperar del enorme esfuerzo a ella dedicado por más de ciento cincuenta científicos españoles y extranjeros, que han trabajado durante dos años con los métodos analíticos más refinados, sigue siendo necesario revisar con rigor crítico la posibilidad infecciosa, al menos para que, en un asunto que abunda en incertidumbres, pudiese tenerse la certeza negativa de que el agente inductor no fue ningún virus o bacteria.

Por si acaso el anterior se considera poco significativo, señalemos otro ejemplo. El predominio de la hipótesis tóxica se basa en "abundantes datos epidemiológicos". Pues bien, hasta ahora no ha habido ninguna exposición sistemática de estos datos y mucho menos una revisión crítica. Se ha convertido en tópica la referencia a la insuficiencia de los trabajos epidemiológicos, pero ignoramos en qué consiste exactamente y qué se está haciendo para remediarla. Los esfuerzos que hemos realizado ante los órganos superiores del Plan Nacional del Síndrome Tóxico para informarnos sobre este y otros extremos han resultado infructuosos ante la increíble cerrazón burocrática de este organismo, rayana en la descortesía. Muchos informes y datos originados en organismos públicos, como son los hospitales, se manejan como si fuesen de propiedad particular, por lo que sólo son accesibles a investigadores que mantienen una relación privilegiada con los centros de origen, quienes los utilizan en sus trabajos o los silencian según su mejor criterio y en las ocasiones que les parecen más oportunas. Muchos investigadores que han tomado parte activa en los trabajos relacionados con el síndrome tóxico se quejan de esta dificultad para obtener información a pesar de repetidas peticiones al respecto.

En segundo lugar, la naturaleza de los datos. De nuevo vamos a limitarnos a unas cuantas muestras. En una enfer-

medad para cuyo estudio tanta importancia han tenido los argumentos estadísticos sorprende la baja calidad aparente de muchos datos y la falta de discusión sobre su validez argumentativa. Hay publicada una distribución de la enfermedad por grupos de edad y sexo en la que no consta la edad de casi el 28 por ciento de la muestra. Entre lo publicado sobre las encuestas epidemiológicas pueden encontrarse afirmaciones de que media una correlación positiva entre el nivel de consumo de aceite en las familias y el riesgo de enfermedad así como las contrarias de que no existe tal relación o, incluso, de que la relación es negativa. En otro orden de cosas, la hipótesis más elaborada de que se dispone actualmente sobre el mecanismo de la enfermedad, basada en modelos experimentales con animales, atribuye una importancia primordial al papel del sistema inmunitario del organismo afectado. Pero los estudios inmunológicos de los enfermos indicaron en su día que "no existe, sin embargo, base objetiva para asegurar que un mecanismo inmunopatológico haya sido decisivo en la producción de ninguna de las lesiones o síntomas característicos de la enfermedad". Tan flagrantes contradicciones son, por supuesto, incompatibles con el pensamiento científico, pero también con el más elemental sentido común, y su eliminación es una tarea urgente.

En fin, la misma dificultad del problema. Una enfermedad nueva, de perfiles cambiantes a lo largo del tiempo, que afecta a muchos sistemas del organismo, rebelde a los tratamientos usuales, producida por un agente elusivo e indetectado, que afecta a unas 20.000 personas y más a las clases bajas, a las mujeres y a los adolescentes, que se produce en un país nervioso por las circunstancias políticas, sociales y económicas, que se asocia a un gran fraude alimentario y que, consciente o inconscientemente, es instrumentalizada desde muy pronto por diversos grupos sociales y políticos, no constituye desde luego el objeto ideal de estudio riguroso. Sólo un enfoque multidisciplinar pero coherente y perseguido con tenacidad, muy exigente respecto a la calidad de los datos, a la limpieza del razonamiento y a la validez teórica y empírica de las conclusiones tendría posibilidades de éxito en tal empeño. Lo realizado en el caso del síndrome tóxico no responde, por desgracia, a este modelo. Muchos y competentes investigadores españoles han realizado grandes esfuerzos y han producido trabajos inobjetables y valiosos sobre aspectos muy especializados del problema. Pero esto no basta. La verdadera comprensión científica no consiste en un amontonamiento indiscriminado de conocimientos fragmentarios, por muy empíricos que sean, sino en conseguir una trama conceptual completa, coherente y fundamentada que dé razón de las causas y los procesos conducentes a lo que se trata de explicar.

A lo largo de estos dos años se ha acumulado un aluvión de datos clínicos, químicos y bioquímicos sobre la enfermedad y sus posibles causas y mecanismos, pero que, tal como están presentados actualmente, no acaban de encajar unos con otros y, además, se hallan huérfanos de un esquema de argumentación más general y de algunas comprobaciones tanto negativas como positivas (por ejemplo, la identificación inequívoca del aceite realmente tóxico, su distribución efectiva y su asociación con los casos reales de enfermedad, tarea sin duda policíaca, pero por la que los investigadores científicos no parecen haberse preocupado mucho o, si lo han hecho, no han debido tener mucho éxito), sin las cuales no acaban de encontrar los cimientos en que apoyarse. La consecuencia es que cuando se intenta dar una visión algo global del problema tengan que abundar las hipótesis "ad hoc" como único medio de colmar las lagunas de conocimiento probado existentes.

Cuando iniciamos la redacción de esta nota nos proponíamos sintetizar lo que actualmente pudiese considerarse conocimiento científico sobre el síndrome tóxico, haciendo algunas observaciones metodológicas y epistemológicas sobre su validez. El estudio de prácticamente toda la documentación disponible y su análisis y discusiones con muchos investigadores españoles aconsejaron cambiar el propósito originario. Lo que ahora nos parece más urgente es poner de manifiesto el derecho de la sociedad española a exigir del Gobierno de la nación explicaciones periódicas de carácter global y coherente sobre la compleja y dolorosa realidad del síndrome tóxico. Ante la confusión reinante se necesitan dos cosas, una de carácter intelectual y otra de carácter moral. La primera es completar, ordenar y depurar la agobiante y a la vez insuficiente información disponible.

Compete a la comunidad científica española realizar la parte fundamental de este trabajo. Es un reto y una obligación, pero también una gran oportunidad. Pocas veces habrá estado la sociedad española tan atenta y receptiva a lo que se hace en los laboratorios y a lo que los investigadores tengan que decir. Será muy importante no desaprovechar esta actitud y dar pruebas tangibles de que la ciencia cumple una función imprescindible en la sociedad. Que, hoy por hoy, no puede darse una explicación definitiva e inobjetable del problema; de acuerdo. Quizá no pueda darse nunca, o quizá se tarden muchos años, como ha sucedido con algunos enigmas sanitarios parecidos. Muy bien. Pero sí puede darse la mejor explicación que haya en cada momento. Hay que arriesgarse a dar esa interpretación verosímil y contrastada y a hacer comprender los obstáculos que impiden conseguir otra mejor. Hemos comprobado que existen los elementos básicos de esta argumentación. Pero están en la mente de cada investigador y no fructifican. Hay que superar la cómoda seguridad de la especialización, si no se quiere traicionar la razón última de ser de la ciencia: comprender la realidad. Y no podrá transmitirse al resto de la sociedad una explicación adecuada, por mínima que sea, si previamente no se la articula dentro de la propia comunidad científica y si no alcanza dentro de ella un cierto grado de aceptación general.

El Gobierno tiene la obligación de hacer que esto sea posible y de exigirlo. Y tras ello tiene la obligación de asumir este nivel de conocimiento que se vaya logrando en cada caso y ofrecérselo a la sociedad revestido de su autoridad moral e institucional. Muchas de las dificultades, quizás insalvables, que se oponen a la resolución del problema provienen del desorden que se produjo en las actuaciones iniciales. Faltó autoridad y eficacia en muchos órganos de la Administración del Estado. Pero al menos el Gobierno asumió el problema y dio unas primeras explicaciones. Luego ha habido, junto a un profundo silencio, un continuo e insensible deslizamiento de esta responsabilidad informativa, que ya en 1982 había pasado a los grupos de investigación concretos y en 1983 parece estar transferida a un organismo internacional como la Organización Mundial de la Salud. Aplaudimos la máxima colaboración de España con estas instituciones internacionales. Pero la responsabilidad y la autoridad necesarias para explicar nuestros problemas públicos a la sociedad española no las tiene más que el Gobierno de la nación. (Francisco Gracia.)

Fuentes termales en el fondo del océano

Expresión de la actividad volcánica que crea nueva corteza oceánica, alimentan extrañas formas de vida y depositan grandes acúmulos metálicos. Contribuyen a explicar la composición química del mar

John M. Edmond y Karen Von Damm

Las fuentes termales son frecuentes en el fondo del océano, a lo largo de las dorsales submarinas, donde tiene lugar la separación de grandes extensiones de la superficie de la Tierra y se crea nueva corteza oceánica. Su existencia era predecible una vez formulada la teoría de la tectónica de placas, que explica el significado de las dorsales; de hecho, J. W. Elder, de la Universidad de Manchester, la postuló hace tiempo. Hoy día, dos décadas después, esta suposición se está confirmando plenamente. Cámaras, sensores y submarinos de investigación que operan a profundidades oceánicas próximas a los tres kilómetros, detectan gran número de fuentes termales; se está comprobando que el agua que aportan al mar tiene importantes repercusiones, de distinta índole.

En primer lugar se observa que la composición química del agua que emana de estas fuentes, y que es el resultado de un complejo conjunto de reacciones entre el agua del mar y las rocas corticales, recién formadas y calientes, alimenta una cadena de organismos que va desde bacterias hasta bivalvos y gusanos tubícolas gigantes. En segundo lugar, parece que las reacciones hidrotermales son la fuente principal de los sedimentos ricos en metales y de los nódulos que tapizan el suelo oceánico. La explotación de estos depósitos es, actualmente, objeto de debate en todo el mundo. Se está demostrando también que otros depósitos metálicos localizados sobre los continentes, y que vienen siendo explotados desde la antigüedad, deben su posición actual a la dinámica de las placas tectónicas, con posterioridad a su formación bajo el mar por efecto de la actividad hidrotermal. En tercer lugar, se confirma que el contenido químico del agua surgente constituye una importante aportación a la composición química de los mismos océanos.

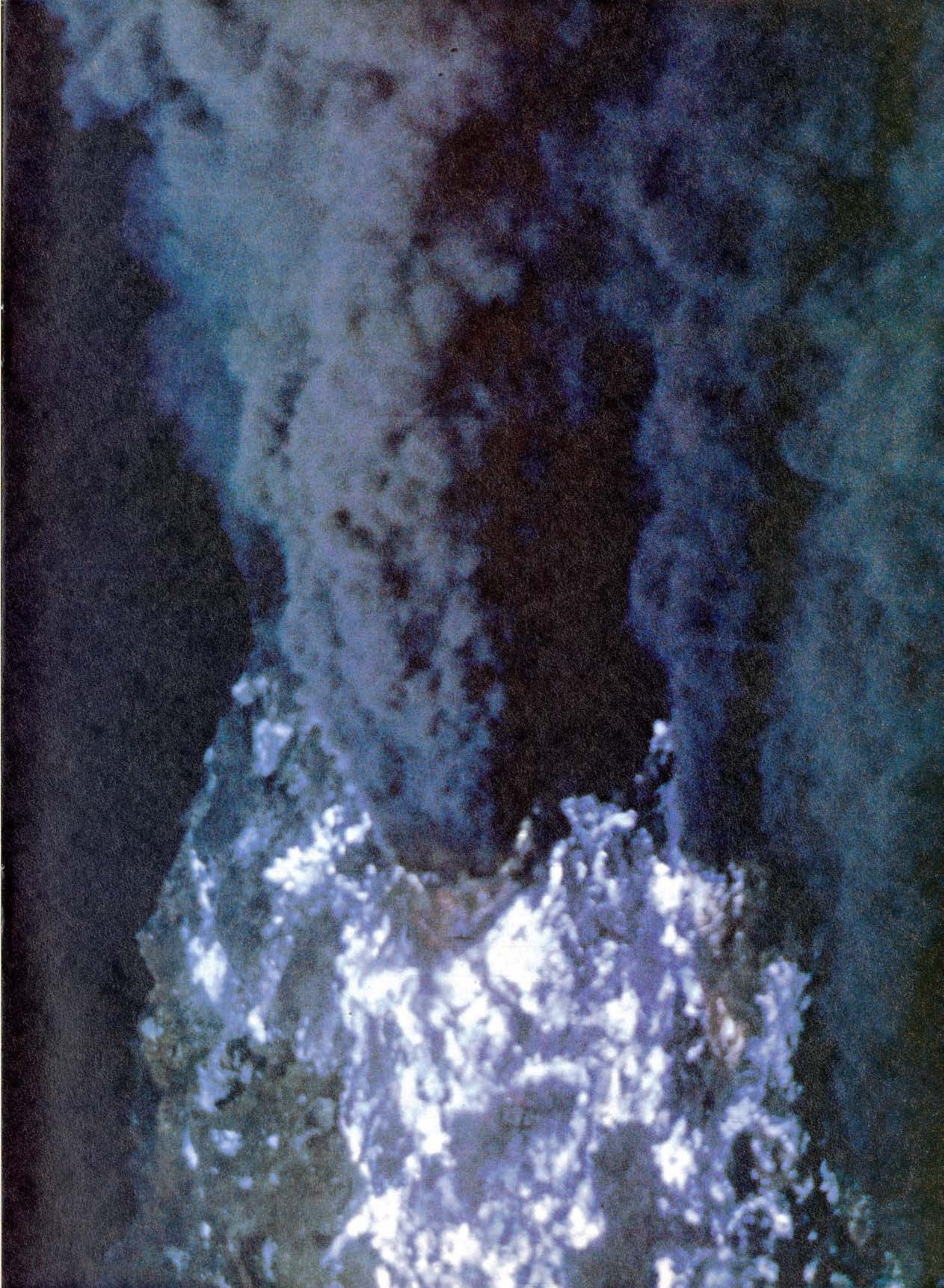
La situación de las fuentes termales resulta bastante clara de acuerdo con la teoría de la tectónica de placas. En dicha teoría la superficie terrestre se supone formada por un conjunto de placas, grandes y rígidas, que se mueven permanentemente unas con respecto a las otras. Las placas están constituidas en su mayor parte por corteza subyacente a los océanos, con grandes retazos de material más ligero, los continentes, inmersos en ellas. A lo largo de los bordes de colisión entre las placas, la corteza antigua se destruye. Son las fosas marinas, lugares donde las placas se sumergen en el interior plástico del manto terrestre. A su vez, en aquellos bordes donde las placas se separan, tiene lugar la formación de corteza nueva. Estos últimos se encuentran casi en su totalidad bajo los océanos y son estructuras lineales de unos 100 kilómetros de longitud, cortadas en zigzag por fallas de transformación de varios kilómetros de recorrido.

Dorsales mesoceánicas

El magma (roca fundida producida por la fusión parcial del manto a profundidades que no exceden escasos centenares de kilómetros) asciende en las dorsales. Su temperatura al alcanzar la superficie es de unos 1200 grados Celsius; se enfría y solidifica, formando corteza oceánica nueva. A su vez, y a medida que progresa el enfriamiento, la corteza se separa lentamente, dejando así espacio para nuevas intrusiones. Este proceso se pone de manifiesto en la propia altura del fondo oceánico. El material caliente se expande; de ahí que las zonas de intrusión sean elevadas. Se trata de las dorsales mesoceánicas, cuyas cimas se sitúan a profundidades de 2,5 a tres kilómetros bajo el mar. Por otro lado, el enfriamiento contrae los materiales, y de aquí que el fondo del océano subsida a medida que se aleja de las dorsales. En consecuencia, la mayor parte del fondo oceánico (y por tanto la mayor parte de la superficie de la Tierra), obedece a una relación gratificadoramente elemental: su altitud decrece con la raíz cuadrada de su edad. En el Atlántico, la parte del fondo que tiene 100 millones de años de edad ha alcanzado su posición actual a casi 1000 kilómetros de la dorsal mesoatlántica, donde se formó, desplazándose a una velocidad aproximada de un centímetro al año. Al enfriarse, ha ido subsidiendo unos 2000 metros, alcanzando así la profundidad del océano en dicha zona los cinco kilómetros.

Era fácil imaginar que, al ser los centros de separación tectónica de las dorsales mesoceánicas zonas de actividad volcánica, debería haber en ellos abundantes fuentes termales, que son prueba de la actividad hidrotermal, como ocurre en el Parque Nacional de Yellowstone y otras áreas de actividad volcánica continental. No obstante, se re-

1. HUMERO NEGRO de una dorsal mesoceánica. Constituye la expresión más espectacular de las fuentes termales y de la actividad hidrotermal del fondo oceánico. El humero se fotografió, desde el submarino *Alvin*, a una profundidad de 2,6 kilómetros, en el Pacífico, justo al sur de la entrada al golfo de California. Se sitúa sobre la cresta de la dorsal del Pacífico oriental. Mide metro y medio de diámetro. La temperatura del agua surgente es de 350 grados Celsius. El agua se infiltra en la corteza oceánica y reacciona con rocas basálticas calientes en una zona de actividad volcánica subyacente a la dorsal, retornando hacia el suelo oceánico como disolución hidrotermal, esto es, un líquido ácido y rico en metales. En su ascenso, la disolución se enfría por mezcla con agua marina y se oscurece por precipitación de partículas de sulfuro de hierro. La chimenea misma la constituyen minerales producidos por precipitación.



querían pruebas de su existencia. No se tardó mucho en lograrlas.

En 1966, Kurt G. T. Boström y Melvin N. A. Peterson, de la Institución Oceanográfica Scripps, analizaron muestras de sedimentos extraídas a ambos lados de las dorsales mesoceánicas. Descubrieron que los centros de separación oceánica, recientemente identificados como tales, estaban cubiertos por óxidos de hierro, manganeso y otros metales; propusieron que dichos sedimentos procedían de la reacción hidrotermal entre el agua marina y la joven corteza caliente. El trabajo subsiguiente, en el marco del *Deep Sea Drilling Project*, dirigido por la Institución Scripps, demostró que la presencia de sedimentos metalíferos constituía un hecho generalizado en la base de la columna de sedimentos oceánicos suprayacente a la roca volcánica que forma la corteza oceánica.

La tectónica de placas, por su parte, colocaba en el primer plano de la atención de los geólogos la roca que forma la corteza oceánica. Se trata del basalto, una roca volcánica de color negro. Químicamente es un silicato alumínico ($AlSiO_3$), sin cuarzo (SiO_2) aunque rico en hierro y manganeso (en contraposición, el granito y la riolita, rocas volcánicas típicas de los continentes, son silicatos alumínicos claros, pobres en hierro y manganeso; constituyen, de hecho, la escoria que ha ascendido hasta la superficie del inmenso horno de fusión que es la Tierra).

En la década de 1960 se realizaron varias campañas de drenaje para muestrear afloramientos próximos al eje de las dorsales. El *Deep Sea Drilling Project* extrajo pequeños testigos de roca procedentes del fondo de los sondeos. El análisis de esas muestras puso de manifiesto que, por lo general, y una vez emplazadas como corteza oceánica, habían experimentado, por lo menos, un episodio de reacción química con el agua del mar. Por ejemplo, durante su enfriamiento hasta unos 500 grados Celsius, las rocas se agrietaron; el agua penetró evidentemente en las fracturas y provocó reacciones químicas y rellenos de las mismas con minerales precipitados, o secundarios.

El proceso se confirmó con datos isotópicos. En cualquier reacción química en que interviene el oxígeno, sus isótopos, o especies nucleares, se fraccionan, esto es, se reparten de manera algo desigual entre los productos de la reacción. La razón fundamental radica en que la energía de un enlace químico depende de las masas de los átomos unidos por el mismo. El grado de fraccionamiento es característico de cada reacción, aunque depende también de la temperatura. El análisis de los mine-

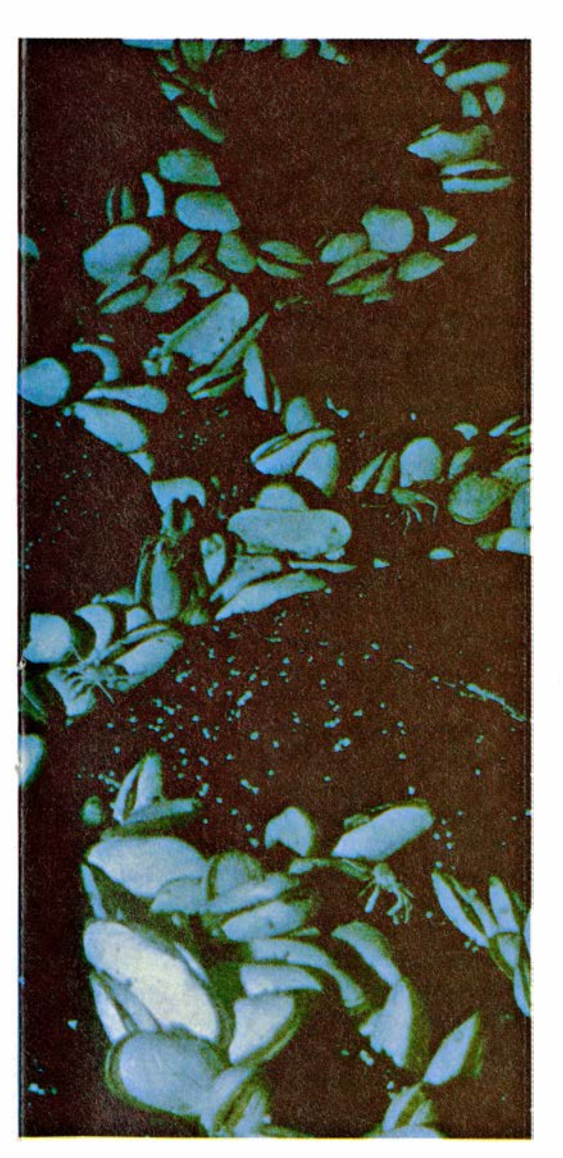


2. BIVALVOS GIGANTES EN UN CAMPO DE FUENTES TERMALES sito al sur del golfo de California. Forman racimos entre los relieves almohadillados del basalto, en las fisuras por donde surgen disoluciones hidrotermales a una temperatura de unos 17 grados Celsius. El agua es rica en sulfuro de

rales secundarios de las muestras de basaltos procedentes de los centros de expansión oceánica confirmó que la composición de sus isótopos de oxígeno difería de la del basalto. Los análisis indicaban que los minerales se habían formado por reacción entre el basalto y el agua del mar a temperaturas no superiores a los 350 grados Celsius.

Ofiolitas

Paradójicamente, fue en los continentes donde se hallaron nuevas pruebas de la actividad hidrotermal de los centros de expansión oceánica. Se han identificado como fragmentos de corteza oceánica, desplazados sobre los continentes por colisiones entre placas, grandes planchones de rocas basálticas que reposan sobre corteza de tipo continental en la isla de Chipre, en el litoral nororiental mediterráneo, en Omán, en el mar Arábigo, en California, en el Tibet y en otros lugares. Llamadas ofiolitas, estas formaciones han permitido estudiar con detenimiento los mecanismos de creación del fondo marino. En estos terrenos ofiolíticos se podía fácilmente muestrear y cartografiar, a lo largo de centenares de kilómetros en horizontal y hasta 10 kilómetros en vertical, lo que otrora había sido corteza oceánica. Las pruebas de actividad hidrotermal a altas temperaturas y a gran escala eran también aquí impresionantes. En el gran complejo ofiolítico de Omán, por ejemplo, el agua del mar había penetrado, sin duda, hasta profundidades superiores a los cinco kilómetros y había reaccionado íntimamente con la roca a temperaturas de unos 400 grados Celsius.



hidrógeno, del que se alimentan ciertas bacterias, que a su vez constituyen el sustento de los bivalvos.

En la mayoría de las ofiolitas, los basaltos están recubiertos por un sedimento metalífero, llamado ámbar, de varios metros de espesor, de aspecto semejante a los que Boström y Peterson hallaron sobre el fondo oceánico actual. Además, abundan en las ofiolitas depósitos metálicos lenticulares que contienen millones de toneladas de sulfuro de hierro (FeS_2, conocido como pirita u "oro de los tontos"). Ocupan depresiones de la superficie del basalto, suprayacentes sobre "chimeneas" de cuarzo y minerales metálicos. Evidentemente, esas "chimeneas", o humeros, fueron en un tiempo los conductos por donde circularon las disoluciones hidrotermales ascendentes, de las que precipitaron el cuarzo, los minerales metálicos y los depósitos lenticulares.

Para determinar directamente el balance calórico de la corteza oceánica a medida que ésta evoluciona, se han realizado, en muchos puntos de los océanos, mediciones de las pérdidas de calor conductivo a través del suelo marino. La pérdida por conducción representa el flujo de calor a través de un medio estático, en este caso, el sedimento que cubre el fondo oceánico. Dichas pérdidas suelen disminuir hacia las dorsales mesoceánicas; sin embargo, la corteza de las dorsales debe estar necesariamente más caliente que la corteza más antigua, alejada de las mismas. De ello se deduce que buena parte del calor transportado hasta las dorsales por el magma ascendente se disipa a través de otro mecanismo, la convección (el transporte de calor mediante la intervención de un fluido, en este caso el agua marina). Efectivamente, Clive R. B. Lister, de la Universidad de Washington, ha deducido que los procesos térmicos que se dan en el eje de las dorsales han de estar dominados por la convección.

La integración, a escala global, de la diferencia entre la pérdida de calor medida y la que debiera esperarse en caso de que dicha pérdida fuera exclusivamente conductiva, arroja un valor cercano a 5×10^{19} calorías por año, casi una décima parte del total del flujo térmico procedente del interior de la Tierra. Esa cantidad debe atribuirse al transporte convectivo. Supongamos que, de acuerdo con los datos isotópicos, la temperatura de funcionamiento de las células convectivas fuera de 350 grados Celsius. Entonces, cada ocho millones de años, aproximadamente, debería circular a través de las dorsales, alcanzando dicha temperatura, un volumen de agua marina equivalente al de los océanos ($1{,}37 \times 10^{21}$ litros). La velocidad del proceso se cifra en torno al 0,5 por ciento del ritmo con el que los océanos reciben agua de los ríos.

Isótopos de helio

La prueba más convincente de la importancia que hoy tiene la actividad hidrotermal en los océanos posiblemente proceda de las determinaciones de la concentración de helio en la columna de agua oceánica. Ese elemento posee dos isótopos: helio 3 y helio 4. Al formarse la Tierra, ambos isótopos se distribuían en una determinada abundancia relativa. Ahora bien, el helio 4 se genera continuamente en el interior de la Tierra por desintegración radiactiva de los isótopos de larga vida del uranio y del torio. Además, tanto el helio 3 como el helio 4 escapan permanentemente hacia el espacio interplanetario desde la alta atmósfera de la Tierra, a la vez que los inyecta a la atmósfera el viento de partículas procedentes del Sol.

La abundancia relativa de helio 3 a helio 4 de un punto a otro de la Tierra es, por consiguiente, la resultante de un conjunto de procesos. Brian Clarke, de la Universidad McMaster, ha desarrollado un espectrómetro de masas de alta sensibilidad para medir la composición isotópica del helio; le ha servido para analizar el contenido en este elemento del agua que mana de los acuíferos a los pozos. Se supone que la concentración de helio 3 no ha experimentado cambios desde el mismo momento en que el agua entró en el acuífero, siendo así un testimonio del equilibrio de solubilidad del helio, esto es, la cantidad del mismo que pasa, de manera natural, al agua desde la atmósfe-

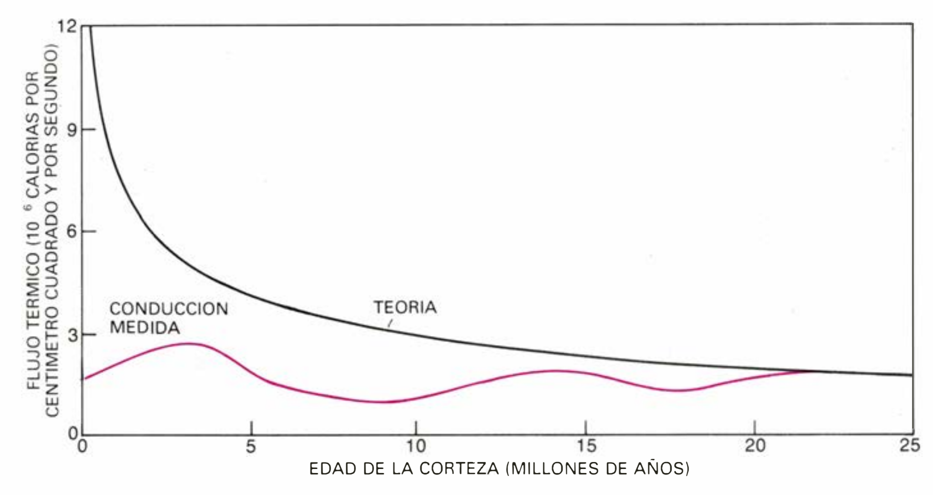
ra. Cualquier anomalía en la concentración de helio 4 respecto a ese valor de equilibrio se considerará resultante de la desintegración radiactiva subterránea. El espectrómetro, por tanto, sirve también para la prospección de uranio.

Clarke, en colaboración con Harmon Craig, de la Scripps, se dispuso entonces a determinar el aporte de helio 4 del manto al océano; como antes, confiaba en que el helio 3 serviría de indicador. Sus hallazgos fueron sorprendentes. En la superficie del Pacífico, el helio disuelto en el agua del mar estaba, efectivamente, en equilibrio con el helio atmosférico. Sin embargo, a profundidades mayores, la cantidad de helio excedía la correspondiente a dicho valor de equilibrio; al sustraer el valor de equilibrio de la concentración medida, la cantidad de helio 3 del exceso era anormalmente alta. La discrepancia alcanzaba un máximo a 2600 metros, profundidad media del eje de la dorsal pacífica. En ella, la relación de helio 3 a helio 4 del helio suplementario era unas ocho veces superior a su valor atmosférico. Evidentemente, el helio 3 quedaba retenido en el manto y se vertía al océano en el eje de las dorsales. ¿Se desprendía el helio en forma de burbujas gaseosas, que se separan de la superficie del magma durante su enfriamiento, o bien se liberaba en el curso de la actividad hidrotermal, responsable de la intensa modificación de las rocas? La magnitud de la anomalía apuntaba hacia la segunda dirección.

¿Cuáles son las reacciones que se producen entre el agua de mar y el basalto del fondo oceánico, a presiones de unos 280 kilogramos por centímetro cuadrado y a temperaturas de varios centenares de grados Celsius? A mediados de los años 70 comenzaron a recogerse los primeros datos, cuando James J. Bischoff y Frank W. Dickson reprodujeron esas condiciones en su laboratorio de la Universidad de Stanford. Los resultados fueron espectaculares. Los iones magnesio del agua de mar se combinaban con el silicato del basalto, rindiendo $Mg(OH)SiO_3$, hidroxisilicato de magnesio, insoluble en agua. El agua aportaba los grupos hidroxilo (OH), enriqueciéndose en iones hidrógeno (H^+); de esta manera, el agua de mar se acidificaba fuertemente. Los iones hidrógeno ocupaban entonces el lugar del calcio y del potasio en la red cristalina del basalto, que en consecuencia recristalizaba.

Composición del agua de mar

En síntesis, las reacciones consistían en un intercambio de iones hidrógeno por otros cationes, o iones positivos, principalmente calcio y potasio. El calcio liberado reaccionaba seguidamente con el sulfato (SO_4) del agua para formar sulfato cálcico ($CaSO_4$), que precipitaba en anhidrita, un mineral. El sulfato podía igualmente combinarse con el hierro del basalto para constituir sulfuro de hierro, esto es, depósitos de pirita. De estas dos reacciones seguíase la extracción completa del sulfato del agua marina. Finalmente, la degradación del basalto y la acidez de la disolución permitían la liberación de sílice (SiO_2 disuelta en el agua).



3. FLUJO TERMICO procedente del fondo del océano, primer indicio de la actividad hidrotermal que se desarrolla en las dorsales mesoceánicas. En éstas, las placas de corteza oceánica se separan, a la vez que asciende nueva corteza, ocupando su lugar; se trata, pues, de zonas notablemente calientes. Sin embargo, las mediciones del calor desprendido por conducción a través de la corteza en tres zonas del Atlántico (*color*) demostró que era muy inferior a la cantidad que debiera esperarse si dicho flujo térmico obedeciera exclusivamente a la conducción (*negro*). La diferencia hay que atribuirla al transporte por convección, esto es, a la circulación de agua a través del fondo oceánico. El eje horizontal muestra la edad de la corteza oceánica, que se correlaciona con la distancia a la dorsal, en la que tuvo lugar su emplazamiento.

¿Contribuirían esas reacciones hidrotermales a la composición del océano en una medida parangonable con la de los productos de la alteración de los continentes? Tal posibilidad resultaba muy atractiva, ya que a mediados de los 70 se había llegado a un punto muerto en el conocimiento de los procesos que controlan la composición del agua marina. En el siglo XIX, antes del florecimiento de la química orgánica que siguió al descubrimiento de los tintes sintéticos, muchos químicos habían emprendido el análisis de las sustancias naturales inorgánicas.

Les animaba a ello la búsqueda de elementos químicos desconocidos que completasen los huecos existentes en el sistema periódico, por entonces en elaboración. El método principal de trabajo era el análisis de los minerales metálicos. R. W. Bunsen, por ejemplo, analizó centenares de minerales, lo que le llevó a interesarse por los propios procesos magmáticos y a viajar a Islandia para observar las erupciones volcánicas. Otros investigadores dirigieron su atención al estudio de las aguas naturales. Así, J. H. van't Hoff sentó las bases de la geoquímica experimental al estudiar la secuencia de precipitados que se formaban a medida que se evaporaba el agua del mar.

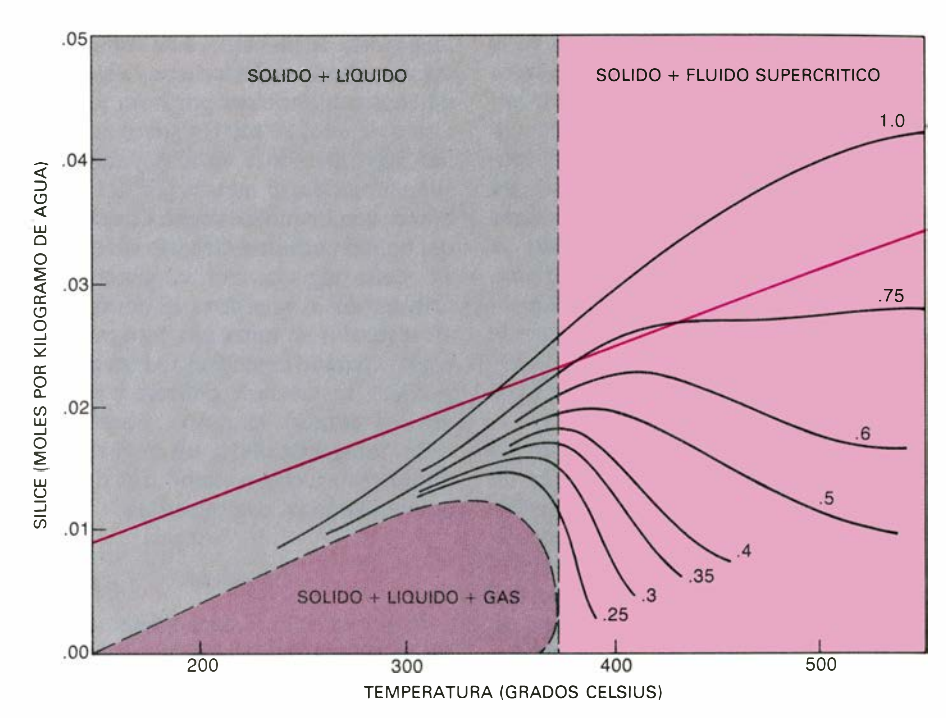
Esta serie de trabajos habían llevado, a finales del siglo pasado, al convencimiento de que el agua de mar no podía producirse por la simple evaporación parcial de agua fluvial. Ejemplos terminales de este proceso se encuentran en las grandes cuencas lacustres cerradas, como el mar Muerto o el Gran Lago Salado, que son altamente alcalinas comparadas con el océano. Siguió luego un largo vacío en la investigación, que se prolongó hasta las postrimerías de la década de 1950, cuando Lars Gunnar Sillén, del Real Instituto de Tecnología de Estocolmo, se propuso resolver el problema. Sillén se preguntaba: ¿Qué es lo que controla al *p*H, o concentración de iones hidrógeno, de los océanos? ¿Por qué su valor se sitúa sistemáticamente entre 7,5 y 8, esto es, bastante próximo a la neutralidad ácido-básica? Acto seguido, Frederick T. Mackenzie y Robert M. Garrels, de la Universidad del Noroeste, tradujeron a términos geológicos las cuestiones, un tanto abstractas, planteadas por Sillén.

Consideremos la alteración de las rocas volcánicas de los continentes. La lluvia que cae sobre las rocas es una disolución ácida, ya que el dióxido de carbono atmosférico se disuelve en ella

formando ácido carbónico (H_2CO_3). Por su parte, las rocas son una red de silicato alumínico en la que están embebidos los cationes. El ácido reacciona con esta red, en concreto, el ácido cede iones hidrógeno, dejando atrás iones bicarbonato (HCO_3^-). El agua de lluvia se hace así alcalina. Los iones hidrógeno ocupan, en la red, el lugar de los cationes, que al liberarse pasan a la disolución. La red se rompe, transformándose en minerales arcillosos altamente desordenados, como la caolinita ($Al_2Si_2O_5(OH)_4$), que se erosionan con facilidad.

Puede estimarse la tasa global de estas reacciones a partir del ritmo de transporte de sus productos al océano por parte de los ríos. El resultado requiere que en unos 4000 años se consuma una cantidad de dióxido de carbono equivalente a su abundancia en la atmósfera. Lógicamente, debe existir una reacción que reconvierta el bicarbonato en dióxido de carbono. Mackenzie y Garrels observaron que la relación de cationes a iones hidrógeno era, por término medio, de $1{,}2 \times 10^3$ en las aguas continentales. En los océanos es de 6×10^7, cuatro órdenes de magnitud más grande. Por consiguiente, cuando los silicatos alumínicos ricos en iones hidrógeno llegan al océano (donde estos iones son escasos y, los cationes, abundantes), las reacciones que dieron lugar a los mismos han de invertirse. Deberían formarse arcillas ricas en cationes, desprendiéndose así los iones hidrógeno, que se combinarían con el bicarbonato para producir nuevamente agua y dióxido de carbono.

Las consecuencias eran de gran trascendencia. La composición global de la corteza continental se ha mantenido constante durante los últimos 2500 millones de años, al menos; por tanto, la composición de los océanos tendrá que haberse conservado también inalterada a lo largo de ese intervalo. Se emprendieron investigaciones encaminadas a determinar si las supuestas reacciones encargadas de reconstituir los minerales arcillosos se desarrollaban en los océanos. Los resultados fueron decepcionantes. Ni siquiera la interpretación más optimista de los datos alcanzaba la escala necesaria para igualar el balance entre la tierra y el mar. Se entraba así en el período de latencia mencionado antes. Por fortuna, los cálculos provisionales sobre los flujos químicos que debían esperarse en relación con la actividad hidrotermal en los ejes de las dorsales alumbraban una vía de solución para el problema.



4. SOLUBILIDAD DEL CUARZO en el agua surgente muestreada en 1977 en el campo de fuentes termales cercano a las islas Galápagos. Demostró, incluso antes del descubrimiento de las chimeneas negras (ocurrido dos años después), que las disoluciones hidrotermales puras que ascienden al suelo del océano debían tener una temperatura de unos 350 grados Celsius. Las muestras no superaban los 19 grados, pero la concentración de sílice (cuarzo disuelto) aumentaba en las más calientes. La extrapolación de esa tendencia (*color*) interceptaba las curvas que representan agua saturada con cuarzo (*negro*), en puntos claramente por encima de los 300 grados. (Las presiones se expresan en kilobares.) Evidentemente, las muestras de Galápagos representaban soluciones hidrotermales diluidas en agua marina ordinaria. Se indican las regiones en que la mezcla cuarzo-agua es estable como (1) sólido más líquido más gas, (2) sólido más líquido y (3) sólido más fluido supercrítico (una fase que no es ni líquida ni gaseosa).

Ahora bien, una cosa es suponer que abundan las fuentes termales a lo largo de los centros de expansión oceánicos y otra, muy distinta, localizarlas bajo una capa de 2,5 kilómetros de agua, por lo menos. Cuando se llevaron a cabo los primeros intentos, a comienzos de los años 70, el conocimiento que se tenía del fondo oceánico y los medios disponibles para explorarlo eran bastante rudimentarios. Los equipos de sonar en servicio sólo funcionaban para terrenos llanos; y así, ante la abrupta topografía de los ejes de las dorsales, apenas registraban más que una confusa mezcla de ecos. Los equipos instrumentales, arrastrados a poca distancia del fondo por necesidades de la exploración, habían de superar una verdadera carrera de obstáculos en la oscuridad, que menoscababa su integridad, cuando no se perdían a un ritmo alarmante. Pese a ello, el dispositivo más complejo, llamado Deep Tow (Remolque Profundo), de la Institución Scripps, se apuntó descubrimientos de indudable valor.

Primeras exploraciones

El Deep Tow, vehículo remolcado al extremo de un cable telemétrico desde un buque, transportaba cámaras de televisión, sonar, sensores de presión y dispositivos para medir la temperatura y conductividad eléctrica del agua (propiedad esta última que guarda relación con la concentración salina). Durante varios años, este ingenio detectó anomalías térmicas en diferentes puntos de los centros de expansión del Pacífico tropical oriental. En cierta ocasión, el Deep Tow, dotado al efecto de los dispositivos adecuados, permitió a Ray F. Weiss, de la Scripps, extraer una muestra de agua tomada en el interior de una de esas anomalías. Su temperatura difería en menos de 0,1 grados Celsius de la del agua ambiental (que estaba a dos grados). Sin embargo, los datos químicos, incluida la concentración de helio 3, demostraban sin lugar a dudas que el agua era de procedencia hidrotermal. Presumiblemente, el Deep Tow había "sobrevolado" una columna de agua, a 15 o 20 metros por encima de una fuente termal submarina.

A mediados de los años 70, la situación comenzó a cambiar rápidamente. Por un lado, la Marina estadounidense facilitó el acceso de los investigadores a las técnicas que había desarrollado para la cartografía del fondo marino, lo que permitió el empleo rutinario de sistemas de navegación profunda de alta

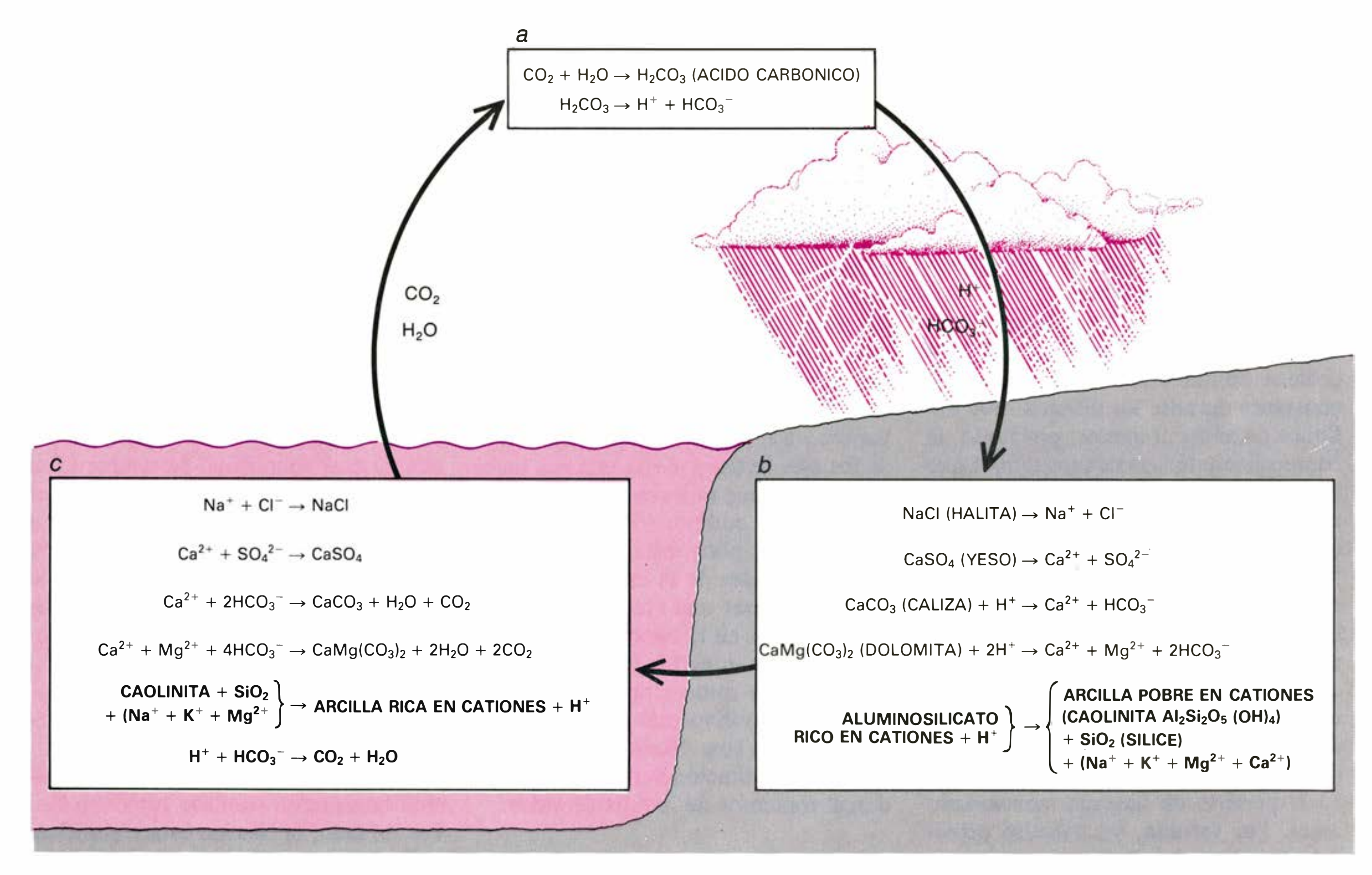
precisión. Tales sistemas se basan en la medida del tiempo que transcurre entre la emisión de un impulso acústico ("impulso pregunta") desde un vehículo submarino y la recepción de las "réplicas" acústicas, procedentes de un conjunto de dispositivos reflectores fijos sobre el fondo. Si previamente se establecen desde un barco, mediante sonar, las posiciones relativas de los reflectores, el vehículo submarino puede dirigirse con bastante facilidad. De hecho, pueden alcanzarse precisiones posicionales de unos pocos metros.

En segundo lugar, se desarrolló en esa época una cámara fotográfica de gran amplitud de campo. Las cámaras disponibles hasta entonces eran instrumentos frágiles y de poco peso, que al ser arrastrados a bastante profundidad bajo el barco resultaban difíciles de controlar. Se requería un vehículo-cámara fotográfica de gran solidez. Y se construyó el Angus, diseñado por la Institución Oceanográfica Woods Hole. Se trata de una "jaula de gorila" de 1,5 toneladas, donde van montados cámaras de color, lámparas estroboscópicas, generadores eléctricos, sonar y reflectores para navegación acústica. En la mayoría de los casos, este vehículo se arrastra desde un barco a una velocidad de cuatro kilómetros por hora y a una altura de unos 20 metros sobre el fondo del mar. Debido a su masa, nunca se sitúa a más de 75 metros por detrás del barco, con lo que los registros del sonar del buque permiten dirigirlo sin que sufra daño alguno. Por lo general, un arrastre del Angus dura 18 horas; cada 10 segundos se toma una fotografía en color. Cuando el vehículo se iza a la superficie, se revela la película y se anotan los detalles extraños. Puesto que cada fotografía lleva un registro del momento en que se tomó, los detalles pueden situarse con exactitud.

El *Alvin*

El tercer avance tecnológico crucial fue el submarino de investigación. Los submarinos que pueden operar a las profundidades de los ejes de las dorsales son pequeños. El mayor de ellos, el *Alvin*, empleado por la Woods Hole, pesa 16,5 toneladas. Consiste en una esfera de titanio resistente a la presión, capaz de albergar a dos investigadores y el piloto. Sin embargo, su reserva energética es muy limitada, lo que explica su lento avance: desarrolla una velocidad máxima sobre el fondo de unos cuatro kilómetros por hora. Además, el alcance de sus luces no sobrepasa los 15 metros. Se trata, pues, de un instrumento poco eficaz para la exploración, cuya aplicación principal es la de visitar puntos seleccionados con anterioridad (a través de fotografías tomadas por el Angus, por ejemplo) y que requieren estudio. El *Alvin* se dirige hacia estos objetivos por medio de las mismas referencias de navegación que emplea el Angus. El *Alvin* puede posarse sobre el fondo a escasas decenas de metros del objetivo escogido.

Esta combinación, perfeccionada y costosa, de tres ingenios se empleó por vez primera en la primavera de 1977, en una dorsal del Pacífico situada a 280 kilómetros al noreste de las islas Galápagos. El buque de investigación *Knorr*, que zarpó de la Woods Hole, fue el primero en llegar; instaló una malla de reflectores sobre el fondo y registró su posición. Comenzó entonces el Angus sus recorridos, bajo la dirección de Robert D. Ballard, de la Woods Hole. Cuando, días más tarde,



5. BALANCE QUIMICO entre la atmósfera, los continentes y los océanos. Lo postuló, por los años 50, Lars Gunnar Sillén, del Real Instituto de Tecnología de Estocolmo, y lo desarrollaron Frederick T. Mackenzie y Robert M. Garrels, de la Universidad del Noroeste. En la atmósfera (*a*) el dióxido de carbono se disuelve en el vapor de agua, formando ácido carbónico. Al caer el agua en forma de lluvia o de nieve (*b*), los iones hidrógeno de la disolución ácida captan los cationes (iones positivos) de las rocas continentales. Los ríos llevan los productos de la alteración a los océanos (*c*). En éstos, dada su pobreza en hidrógeno, se invierten las reacciones, formándose rocas sedimentarias ricas en cationes y retornando el dióxido de carbono a la atmósfera.

llegó el *Alvin*, a bordo de su pequeño barco nodriza, *Lulu*, ya se habían identificado varios objetivos. En todos ellos, las fotografías del Angus mostraban algunos bivalvos blancos de gran tamaño sobre un fondo negro de basalto. Se condujo el *Alvin* hacia uno de esos puntos; formaban el equipo investigador de a bordo John B. Corliss, de la Universidad estatal de Oregón, y uno de nosotros (Edmond).

El *Alvin* se dejaba sumergir pasivamente a una velocidad de 30 a 35 metros por minuto. Al cabo de hora y media había alcanzado una profundidad de 2500 metros. En ese punto, a unos 100 metros del fondo, soltamos lastre para adquirir una flotabilidad neutra. Seguimos luego descendiendo hasta casi tocar el fondo del mar, que resultó ser una suave pendiente. Durante casi media hora navegamos de un lado para otro buscando el objetivo. Disponíamos cada uno de una claraboya de plexiglás para observar el exterior. Finalmente paramos para recoger algunas muestras de roca. Mientras nuestro piloto empleaba el brazo mecánico del submarino en una ardua maniobra para coger una almohadilla basáltica (estructura volcánica en forma de almohadón, formada, durante la extrusión lenta de lava sobre el fondo marino, al enfriarse bruscamente su superficie), nuestra atención quedó centrada en un par de grandes anémonas marinas de color púrpura. Al girar la vista, nos percatamos de que en el radio de acción de nuestras luces el agua tremulaba, como el aire sobre un pavimento caliente. La temperatura, que medimos apresuradamente, se hallaba unos cinco grados por encima de la del agua ambiental (a 2,05 grados Celsius). Olvidados de las rocas, tomamos una muestra del agua y continuamos el recorrido pendiente arriba. Poco después desembocábamos ante un espectáculo fabuloso.

Un mundo desconocido

El terreno basáltico típico del eje de la dorsal es puro yermo: monótonas extensiones de lavas almohadilladas marrones, interrumpidas por fallas y fisuras. Para encontrar un simple organismo hay que examinar varios metros cuadrados. Y sin embargo, había aquí un verdadero oasis. Arrecifes de mejillones y campos de bivalvos gigantes se bañaban en las trémulas aguas acompañados por cangrejos, anémonas y grandes peces rosa. Lo que restó de las cinco horas de inmersión lo vivimos con un verdadero frenesí. Medimos la temperatura, conductividad, *p*H y contenido en oxígeno del agua; tomamos fotografías; muestreamos el agua y nos aseguramos de que se hubiese recogido una selección representativa de los organismos existentes. Todo ello bajo la creciente amenaza del continuo descenso de los voltajes de nuestro equipo.

Afortunadamente, éste funcionó a la perfección. Pronto nos dimos cuenta de que nos encontrábamos ante un campo de fuentes termales. Chorros de agua caliente brotaban de cualquier orificio o fisura del fondo marino, en un área circular de unos 100 metros de diámetro. La temperatura del agua era variable en un amplio rango, con un valor máximo de 17 grados Celsius. Los organismos se localizaban de manera bastante selectiva, bloqueando preferentemente los orificios más cálidos. En algunos casos los arrecifes de mejillones formaban verdaderos conductos, canalizando de hecho la corriente de agua.

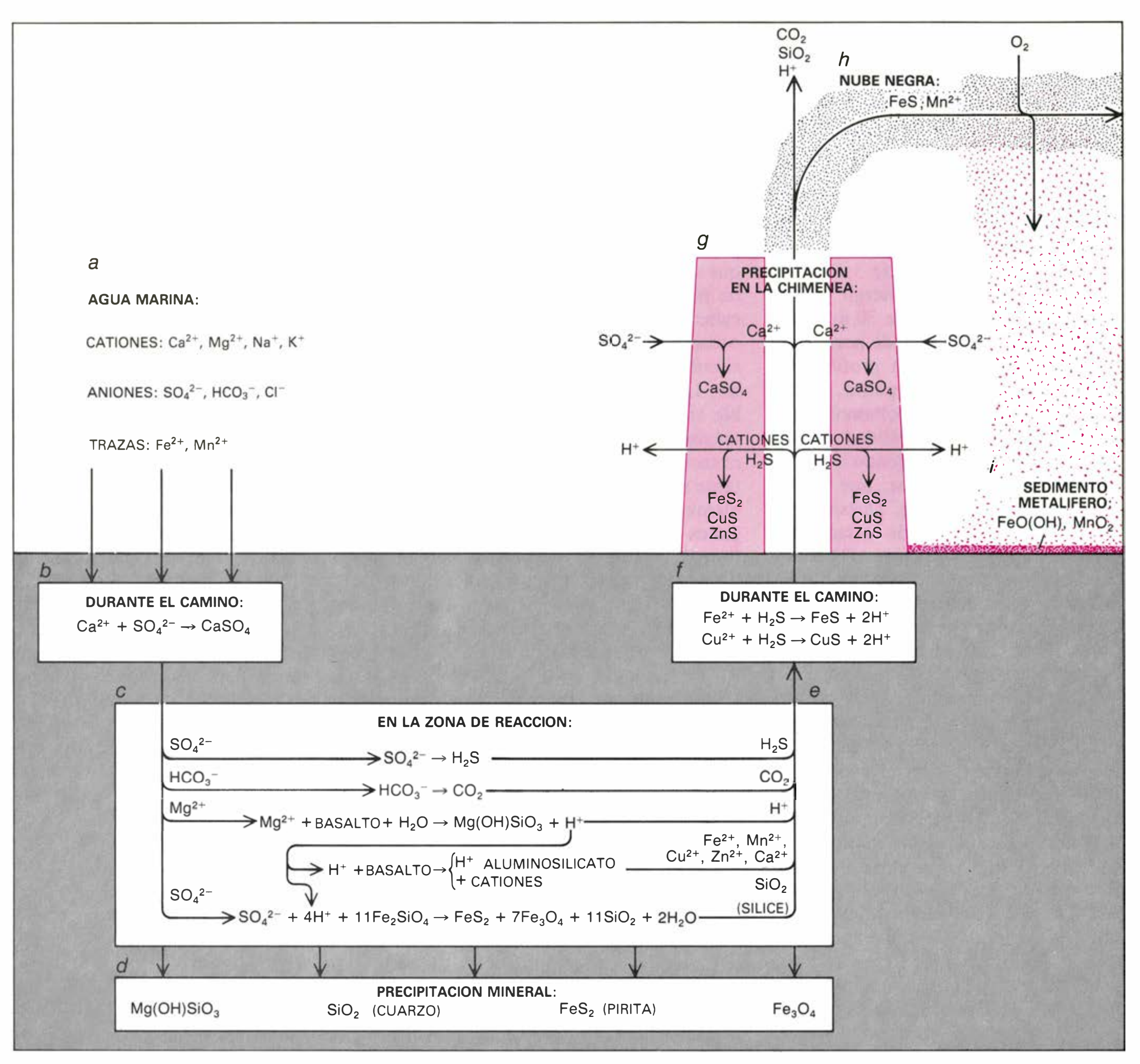
Trabajamos hasta agotar la energía reservada a la investigación científica, llegado a lo cual soltamos más lastre. El *Alvin* abandonó el fondo. Apagamos el equipo y comenzamos a sentir frío. Tras permanecer ocho horas agazapados en esa esfera de dos metros, donde casi cada movimiento requería la cooperación de los otros colegas, a su vez en posiciones no menos incómodas, empleamos los 90 minutos del ascenso en intentar conservar el calor, demasiado cansados para prestar atención a los organismos bioluminiscentes que huían al paso del submarino. La aproximación a la superficie nos pareció, como siempre, un nuevo nacimiento. A una profundidad de unos 200 metros, el agua comenzó a adquirir una luminosidad verde pálido. Poco después todo era claridad. El piloto vació los tanques de lastre, rodeando el *Alvin* con una columna de burbujas, y emergimos con un suave impulso. Salimos a flote, meciéndonos tranquilamente en el oleaje. Inmediatamente los buceadores rodearon el submarino, fijando los cables para su recuperación, y acto seguido se izó el *Alvin* hasta su alojamiento. Quedaba la agonía de tener que ponerse en pie y salir gateando. En la cubierta del *Lulu* todos compartían ya nuestra excitación. La bolsa de muestras del *Alvin* estaba cargada de bivalvos y mejillones. Algunos cangrejos que se habían introducido bajo la envoltura de fibra de cristal del *Alvin* comenzaron a caer sobre la cubierta.

El verdadero trabajo comenzaba ahora. Se desmontaron los recogedores de muestras de agua y se trasladaron éstas al *Knorr* para analizarlas esa misma noche. Se revisaron las fotos del Angus a fin de localizar nuevos objetivos. Se limpiaron los portamuestras de agua y se volvieron a instalar en el submarino. A las seis en punto de la mañana del día siguiente, el submarino *Alvin* estaba listo para una nueva exploración; a las 8:45 iniciaba otro descenso. Inmersión que se repetiría en 15 ocasiones.

Sulfhídrico por alimento

A medida que progresaba el trabajo, se hacía más claro que todas las aguas procedentes de fuente termal tenían un elevado contenido en sulfhídrico (H_2S). Ahí residía la explicación de los oasis submarinos. Efectivamente, las bacterias que derivan su energía de la oxidación del sulfhídrico son corrientes en muchos sistemas ecológicos. Sin embargo, en los oasis, las bacterias tenían que ser los productores primarios, situándose por consiguiente en la base de la pirámide ecológica. La energía que rige estos recién descubiertos ecosistemas no procede del Sol, como en el caso de la fotosíntesis, fuente de energía en cualquier otro ecosistema, sino de la desintegración radiactiva de los isótopos de larga vida del uranio, torio y potasio en el interior de la Tierra. Esa desintegración desprende calor; el calor genera magma que, a su vez, se convierte en nueva corteza oceánica. El agua marina percola en la corteza y reacciona con las rocas a elevada presión y temperatura.

De las distintas reacciones, dos desempeñan un papel fundamental en los oasis. El sulfato disuelto en el agua de mar reacciona con el hierro de la roca, produciendo sulfuro de hidrógeno y óxidos de hierro; asimismo, los minerales de tipo sulfuro contenidos en la roca experimentan disolución. En consecuencia, el agua caliente que asciende de vuelta hacia el fondo del océano tiene un alto contenido en sulfuro. En los puntos del fondo donde brotan esas aguas, el sulfuro de hidrógeno es absorbido por bacterias, que extraen también oxígeno de origen fotosintético disuelto en el agua ambiental. Ambos se combinan, formándose sulfato a partir del sulfuro. La energía que se desprende de esta reacción rige el metabolismo de las bacterias, que, a su vez, sirven de alimento a otras especies. La propia diversidad de éstas era ya, de hecho, una prueba evidente de la abundancia de fuentes termales junto a los ejes de las dorsales en todo el mundo. ¿Cómo, si



6. REACCIONES HIDROTERMALES que se desarrollan en las dorsales mesoceánicas. Regeneran dióxido de carbono en una proporción muy superior a la prevista por Sillén para los procesos oceánicos de baja temperatura. El agua marina (*a*) se infiltra en la corteza oceánica. A medida que desciende, parte de su contenido iónico puede precipitar (*b*) como sulfato cálcico. Los restantes iones prosiguen hasta una zona situada varios kilómetros bajo el fondo. Reaccionan allí con rocas basálticas calientes (*c*) para dar nuevos precipitados minerales (*d*) y una disolución hidrotermal caliente, ácida y rica en metales (*e*), que asciende hacia el suelo marino. Esta disolución ascendente puede encontrarse con agua marina fría, enfriándose a su vez y precipitando sulfuros metálicos (*f*). Al llegar al suelo, la disolución se combina con el sulfato del agua marina, formando chimeneas de precipitados de sulfuros metálicos y sulfato cálcico (*g*). Sobre las chimeneas, el hierro todavía en disolución forma nubes negras de sulfuro de hierro (*h*). El manganeso persiste en disolución. Finalmente, tanto el hierro como el manganeso se oxidan y caen cual lluvia sobre el fondo del océano, generando sedimentos metalíferos, omnipresentes en la corteza oceánica (*i*). Por su parte, el dióxido de carbono surgente se mezcla con el océano; llega a la superficie y retorna a la atmósfera.

no, podría desarrollarse y mantenerse una fauna única y tan altamente evolucionada?

A nuestro laboratorio del Instituto de Tecnología de Massachusetts le correspondió el análisis pormenorizado del agua recogida por el *Alvin* en las surgencias submarinas. En síntesis, la composición química del agua era consecuencia de un complejo conjunto de reacciones entre el agua marina y el basalto, que tenía lugar en una zona inaccesible, probablemente situada a varios kilómetros bajo el fondo del mar. De ahí que la mejor manera de desvelar dichas reacciones fuera determinar el grado de enriquecimiento o empobrecimiento de varios elementos químicos en las muestras de agua, respecto al basalto y al agua de mar ordinaria. Acabamos analizando la concentración de 35 elementos, probablemente la mayor cantidad nunca medida en un esfuerzo de ese tipo. A medida que avanzábamos se iba desvelando la dinámica del proceso químico.

El magnesio y el sulfato, componentes ordinarios del agua de mar, mostraban un descenso uniforme de sus concentraciones al aumentar la temperatura (es decir, la temperatura de la muestra de agua al ser recogida por el *Alvin*). Sin duda las muestras representaban agua surgente diluida en cierta medida por agua de mar. La extrapolación de la curva de concentración-temperatura, suponiendo que el contenido de magnesio y de sulfato en el agua surgente fueran nulos, arrojaba

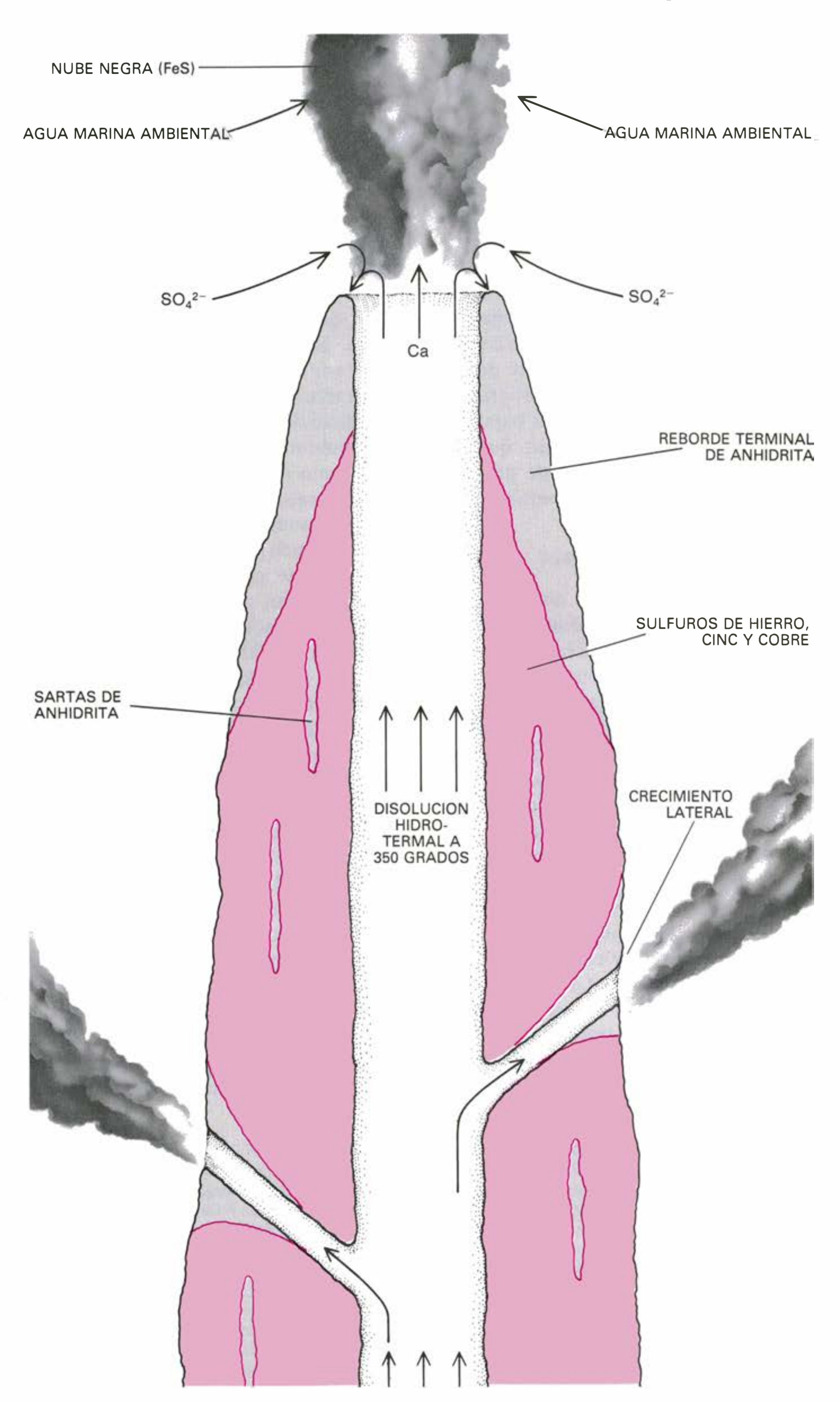
una temperatura en torno a los 350 grados Celsius. Hay que reconocer, sin embargo, lo extravagante de esta extrapolación, pues la temperatura más alta que pudimos medir fue de tan sólo 19 grados. Por otro lado, la concentración de sílice (cuarzo disuelto en agua), aumentaba al hacerlo la temperatura. La extrapolación de esta tendencia hasta su punto de intersección con las curvas que describen la solubilidad del cuarzo en agua, a diversas presiones y temperaturas, daba también una temperatura en torno a los 350 grados Celsius.

Propiedades del agua surgente

El hierro, cobre, níquel y cadmio, que se combinan con el sulfuro de hidrógeno del agua surgente formando precipitados de óxidos y sulfuros, mostraban todos un acentuado descenso al aumentar la temperatura, cuyos valores, al extrapolar la curva a cero, se situaban entre los 30 y los 35 grados Celsius. Tal intervalo podía interpretarse como correspondiente a la temperatura de un reservorio de agua purgado por las fuentes termales. Supusimos así que el agua ascendente desde un régimen de reacciones a alta temperatura entraba en contacto con aguas freáticas cuya temperatura y composición debían estar próximas a las del agua marina ambiental sobre el eje de la dorsal. La disolución de alta temperatura se enfriaba al mezclarse con esas aguas, a la vez que su acidez disminuía, permitiendo la precipitación, en el sistema de conductos hidráulicos del basalto, de los elementos mencionados, en forma de óxidos y de sulfuros. La precipitación debía ser tan extrema, que los elementos que formaban los óxidos y los sulfuros llegaban a extraerse de las propias aguas freáticas. Por tanto, el contenido de esos elementos en nuestras muestras tenía que buscarse en la contribución del agua ambiental en la que se diluyeron. Pese a ello, la prueba de la existencia de reacciones a alta temperatura se conservaba en los datos de magnesio, sulfato y sílice.

El helio 3 aumentaba sensiblemente con la temperatura, a un ritmo de $2{,}2 \times 10^{-17}$ moles (unos 13 millones de átomos) por cada caloría. En el helio remanente, una vez sustraída la cantidad en equilibrio de la concentración medida, la relación de helio 3 a helio 4 era ocho veces superior a la de la atmósfera. Poco antes de descubrir semejante concentración de helio 3 en las profundidades del Pacífico, Craig, Clarke y sus colaboradores habían calculado que la cantidad de helio 3 requerida para mantener la anomalía mundial, habida cuenta de las pérdidas de helio hacia la atmósfera, debía ser de unos 1100 moles por año. Si todo el helio 3 que se añade al océano procede de las fuentes termales de los ejes de las dorsales, y si la relación de helio 3 a calor es, en cualquier punto, la misma que la establecida sobre el eje de la dorsal de las Galápagos, puede deducirse fácilmente el transporte de calor a escala mundial a través de las fuentes termales. Se obtiene un valor de 5×10^{19} calorías por año, que cae en el centro del intervalo estimado.

Podemos considerar, pues, justifica-



7. CRECEN LOS HUMEROS NEGROS a partir del sulfato cálcico que precipita de la disolución hidrotermal y forma un "reborde terminal" del mineral anhidrita. A medida que se expande, el reborde se expone a la acción de la disolución caliente no diluida. Al redisolverse, sin embargo, actúa de molde para la deposición de sulfuros metálicos. La anhidrita puede conservarse si queda englobada por los sulfuros.

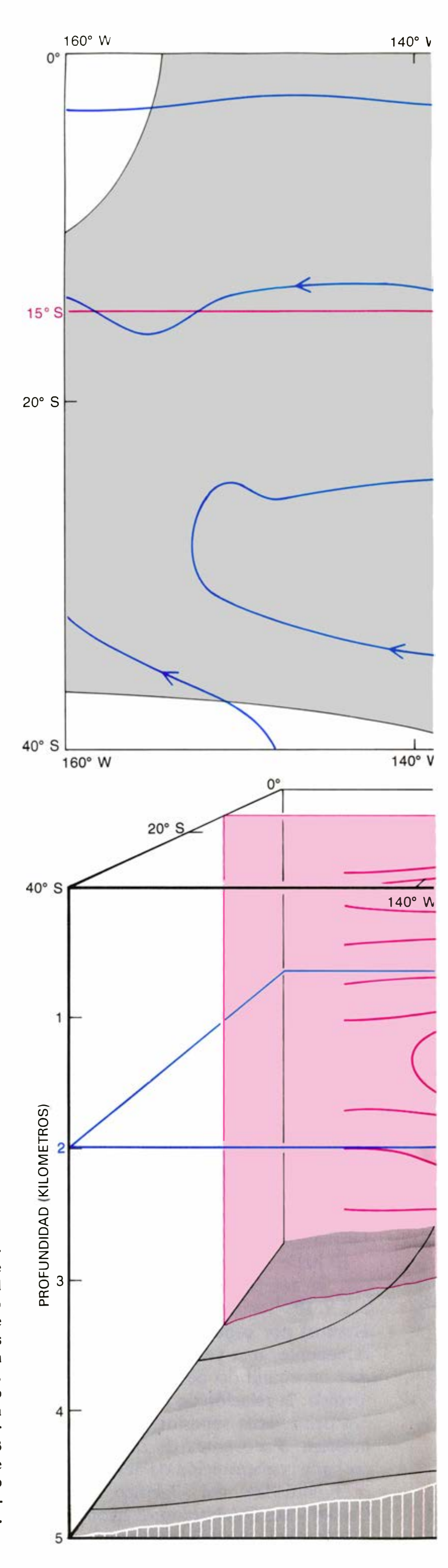
da la suposición de que las anomalías de los restantes elementos en nuestras muestras podían servirnos (referidas a la anomalía de helio 3) para calcular los aportes totales de los mismos al océano a través de la actividad de las fuentes termales submarinas. Los resultados a que se llega son impresionantes, en particular si los comparamos con los aportes procedentes de la alteración continental. En los ejes de las dorsales se consume la mayor parte del magnesio y del sulfato que vierten los ríos al mar, mientras que se desprende entre cinco y 10 veces más litio y rubidio, y entre un tercio y la mitad de la cantidad de potasio, calcio, bario y sílice. Los aportes de manganeso son suficientes para explicar su acumulación en los sedimentos metalíferos y en los nódulos que tapizan el océano. Finalmente, los ejes de las dorsales reconvierten en dióxido de carbono la mayor parte del bicarbonato producido por la alteración continental. Las reacciones a baja temperatura intuidas por Sillén quedan, pues, desplazadas a favor de procesos hidrotermales de alta temperatura.

Humeros negros

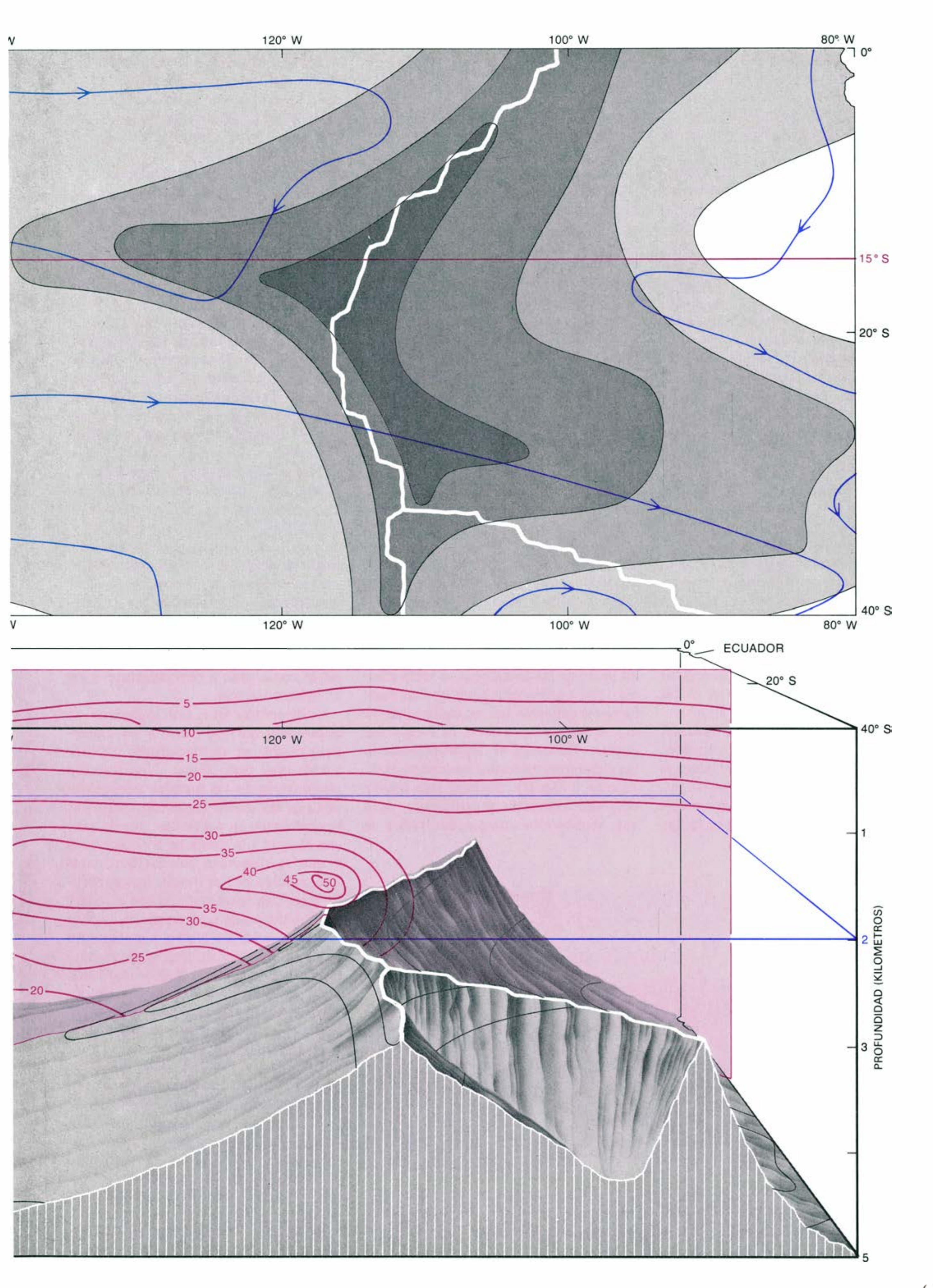
A partir de los datos analíticos, se imponía modificar, desde su base, la interpretación de los procesos que controlan la química del agua de mar a lo largo del tiempo geológico. Por supuesto, lo ideal sería encontrar surgencias de aguas calientes submarinas no afectadas por dilución alguna. La profunda infiltración de agua freática en el eje de la dorsal de Galápagos hacía, sin embargo, altamente improbable un caso así. Por ello, cuando nos enteramos de los hallazgos de Jean Franchetau, del Institut de Physique du Globe, los recibimos con sorpresa y emoción. El submarino oceanográfico *Cyana*, con pabellón francés, había encontrado grandes depósitos de sulfuros junto a la cresta de la dorsal del Pacífico oriental, justo al sur de la entrada del golfo de California, a 21 grados de latitud Norte. El punto no mostraba signos aparentes de actividad hidrotermal. Sin embargo, el depósito sólo podía haberse formado por precipitación de sulfuros a partir de una disolución a alta temperatura.

Volvimos al eje de la dorsal de Galápagos en la primavera de 1979. Se localizaron muchos más campos de surgencias, pero la temperatura del agua en ningún caso superaba los 23 grados Celsius. Finalmente, el *Alvin* partió para intervenir en los trabajos próximos al golfo de California. El Angus, que ya estaba allí, había fotografiado varios campos de fuentes termales y, en su interior, el *Alvin* halló lo que creíamos ilocalizable. Grandes chorros de agua negra ascendían desde humeros de sulfuros, que alcanzaban varios metros de altura. La temperatura del agua surgente superaba los 300 grados Celsius (a esa temperatura, y a la presión de la superficie del mar, el agua ebulliría explosivamente, tal es el caso de los géiseres, como el famoso de Yellowstone; sin embargo, las presiones reinantes a tales profundidades mantienen el agua en fase líquida). Se desencadenó entonces una frenética actividad encaminada a construir un portamuestras de agua capaz de funcionar a temperaturas superiores a los 300 grados. La Fundación Nacional para la Ciencia concedió al *Alvin* más tiempo para trabajar en las "chimenas negras"; John A. Archuleta, del Laboratorio Nacional de Los Alamos, nos prestó el equipo necesario.

Nos sumergimos en noviembre de 1979. Los reflectores de navegación dejados en el lugar con anterioridad nos guiaron hasta las surgencias. Nos acercamos a una de las chimeneas negras, que sobresalía del fondo entre grandes bloques de sulfuros metálicos. Maniobrando en una poderosa corriente de fondo, el piloto insertó un medidor de temperaturas de nuevo diseño en el mismo cuello de la chimenea, de tan sólo 15 centímetros de diámetro. La



8. DISTRIBUCION DE SEDIMENTOS METALIFEROS sobre el fondo del Pacífico meridional, establecida por Kurt G. T. Boström, de la Institución Oceanográfica Scripps. Es asimétrica con respecto al eje de la dorsal de donde surgió el metal en forma de disolución hidrotermal. El dibujo superior representa el Pacífico meridional en planta. A 15 grados de latitud Sur se proyecta hacia el Oeste un largo y estrecho saliente de sedimento, rico en óxidos de hierro y manganeso; se indica con tonos grises. A 30 grados de latitud Sur se proyecta hacia el este un nuevo saliente, más corto y ancho. Las líneas azules señalan las corrientes oceánicas a una profundidad de 2000 metros (a unos 600 metros sobre el eje de la dorsal), según las ha deducido Joseph L. Reid, de la Institución Scripps. Las corrientes han arrastrado a los óxidos, alejándolos del eje; ello explica la presencia de los salientes. El dibujo inferior corresponde al Pacífico meridional visto desde el sur. Las líneas rojas señalan las concentraciones del isótopo helio 3; definen una pluma que se extiende hacia el oeste, a 15 grados de latitud Sur. Las mediciones fueron realizadas por John E. Lupton y Harmon Craig, también de la Scripps. Los números indican la concentración de helio 3 en términos del porcentaje en exceso respecto al valor de saturación normal en el mar. No cabe duda de que el helio se desprende de la corteza oceánica por reacciones hidrotermales. Puesto que es inerte, no experimenta cambios al ser expulsado en los ejes de las dorsales; se comporta, por tanto, como un verdadero marcador del agua oceánica hidrotermal de la que precipitan los sedimentos metalíferos.



120° W
100° W
80° W
0°
15° S
20° S
40° S
ECUADOR
5
10
15
20
25
30
35
40
45
50
1
2
3
PROFUNDIDAD (KILOMETROS)

lectura se estabilizó, a menos de unas décimas de grado, a 350 grados Celsius. Se confirmaba así el poder predictivo de la química. En todas las surgencias muestreadas, la temperatura no se apartó más que unos pocos grados de los 350.

Las disoluciones surgen de las chimeneas, a esas temperaturas, en forma de fluidos claros y homogéneos. Contienen 100 partes por millón de hierro y unas pocas partes por millón de cinc, cobre y níquel, lo que representa un enriquecimiento del orden de 10^8 respecto del agua marina, donde las concentraciones de esos elementos se miden en partes por billón. Además, las disoluciones contienen 210 partes por millón de sulfuro de hidrógeno, componente del que carece el agua marina ordinaria. La concentración de sílice es de 1290 partes por millón, coincidente con lo que se deduce del examen de los datos de las Galápagos. En otras palabras, las disoluciones, saturadas de cuarzo, carecen de magnesio y sulfato, en perfecto acuerdo con la predicción.

En la boca de las surgencias, las disoluciones se mezclan rápidamente con el agua marina ambiental, fría. Se forma así el "humo" negro de la emanación; se trata de una suspensión de finas partículas de sulfuro de hierro que precipitan de la disolución. Las chimeneas parecen crecer, por acreción de sulfato cálcico, en reborde terminal; el calcio procede de las disoluciones hidrotermales y el sulfato lo aporta el agua del mar. A medida que crece el reborde, queda expuesto a la acción de las aguas surgentes, no diluidas; se redisuelve entonces, para ser reemplazado por precipitados de sulfuros. Con frecuencia, las aguas surgentes se escapan a través de los poros de las chimeneas, lo que provoca un crecimiento lateral de las mismas. El resultado es un complejo depósito multimineral.

Explotación de los metales

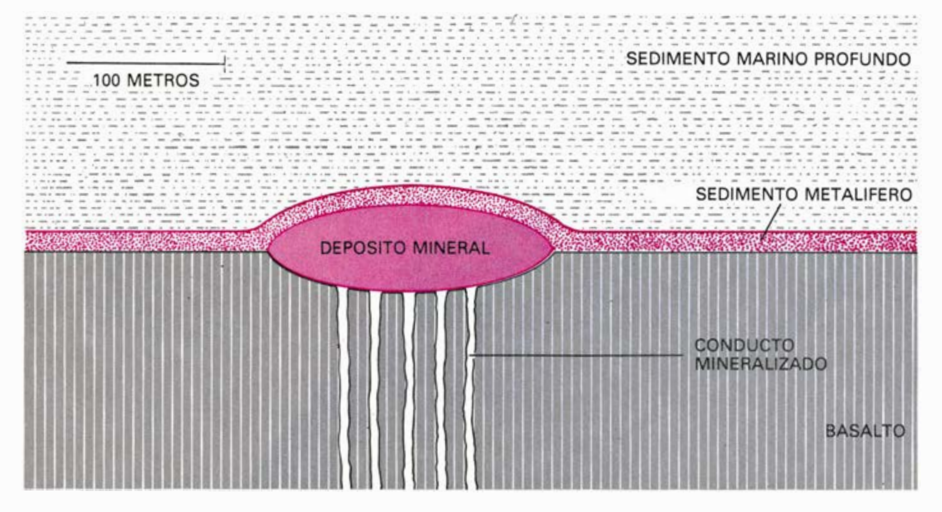
La explotación minera de las disoluciones que ascienden por las chimeneas no merece el esfuerzo que exige. Así, la recuperación de una tonelada de cinc implicaría el procesamiento de casi seis millones de toneladas de disolución, que es la cantidad arrojada por un humero negro típico en un período de varios meses. Puesto que una tonelada de cinc se valora, aproximadamente, en 900 dólares (unas 119.000 pesetas), parece mejor dejar que la naturaleza siga su curso y centrarse en la explotación de los depósitos metálicos resultantes.

Empecemos por las chimeneas negras. La formación de un gran depósito de sulfuros, como los de las ofiolitas, que contienen millones de toneladas de mineral, parece requerir el concurso de un verdadero bosque de chimeneas. Es probable, además, que éstas ocuparan originalmente un área deprimida del fondo oceánico, donde se irían acumulando las partículas precipitadas de las disoluciones surgentes. En cualquier caso, el depósito hidrotermal formado en torno de las chimeneas se vería afectado con posterioridad por nuevas disoluciones calientes que se infiltrarían en él, en su ascenso hacia el fondo del océano. Con ello, el depósito perdería los elementos minoritarios (cobre, cinc, cobalto y níquel), dejando una matriz de grandes cristales de pirita pura. A su vez, dichos elementos minoritarios se redepositarían sobre la pirita, en torno a nuevas chimeneas situadas sobre el fondo del océano. La infiltración de agua marina fría en el depósito en proceso de acumulación favorecería, por su parte, la precipitación "in situ" de depósitos frescos a partir de las disoluciones hidrotermales ascendentes.

Esta secuencia de fenómenos explica las características generales de los depósitos que se explotan en las ofiolitas. La parte principal de cualquiera de ellos consta de pirita pura de grano grueso. La capa suprayacente a la pirita, que se ha oxidado a ocres, está enriquecida en elementos minoritarios. Alexander Malahoff, de la Administración Nacional Oceanográfica y Atmosférica de los Estados Unidos, ha descubierto recientemente una de esas selvas de humeros negros. A unos 30 kilómetros de distancia de los campos hidrotermales de Galápagos observó, desde el *Alvin*, una auténtica profusión de chimeneas extintas en un tramo de unos 600 metros del eje de la dorsal. Sobresalían de un depósito de sulfuros de varias decenas de metros de altura. Dada la cantidad de calor que hubiera emanado de las chimeneas, de haberse mantenido en actividad, es poco probable que se hubiesen podido estudiar de cerca con un submarino de investigación. Las claraboyas de plexiglás del *Alvin* comienzan a reblandecerse a los 86 grados Celsius.

Cabe señalar que, por lo general, la deposición mineral junto a las chimeneas negras es un mecanismo concentrador muy poco eficaz. Una fracción mayoritaria de los metales transportados por las disoluciones hidrotermales se convierten en partículas, que oscurecen el agua a la salida de las chimeneas y que se dispersan con las corrientes que circulan por el fondo. Las partículas terminan reaccionando con el oxígeno disuelto en el agua de mar y, puesto que los óxidos de hierro y manganeso son insolubles, precipitan formando depósitos metalíferos. Tales depósitos contienen pequeñas cantidades de cobre metálico, cinc, cobalto y níquel; pese a ello, son demasiado pobres para su explotación comercial.

¿Dónde se depositan, pues, los sedimentos ricos en metal? A raíz del descubrimiento de Boström y Peterson de que los sedimentos metalíferos se extendían por todo el fondo oceánico, comenzó el primero a elaborar un mapa a escala mundial de distribución del contenido en metales. Sobre el fondo del Pacífico meridional observó un hecho significativo. A unos 15 grados de lati-



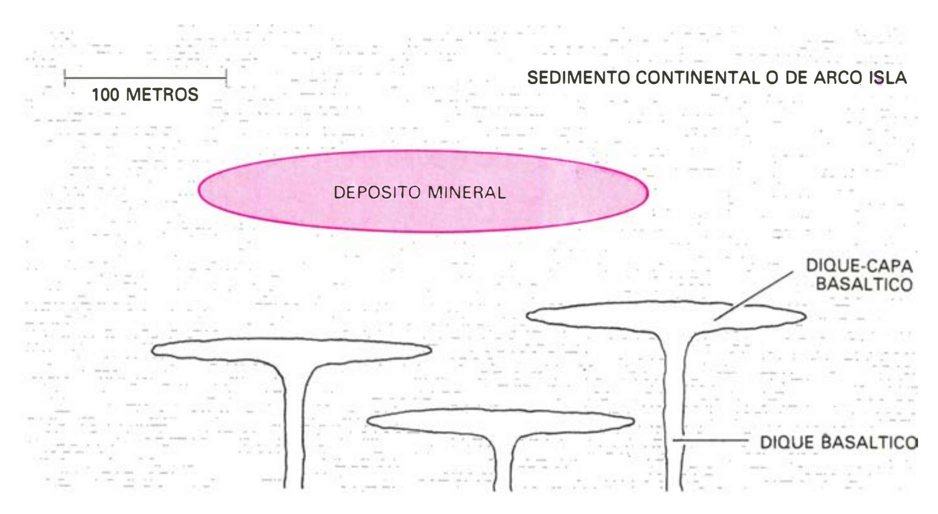
9. UNA OFIOLITA, fragmento de corteza oceánica emplazada sobre un continente, consta de capas de sedimentos idénticos a los que tapizan el fondo oceánico. También contiene acúmulos de sulfuros, probablemente resultantes de la actividad hidrotermal en el fondo oceánico a través de una multitud de humeros negros. A menudo, los depósitos metálicos se localizan sobre depresiones de la capa de basalto.

tud Sur, y proyectándose hacia el Oeste desde el eje de la dorsal, se evidenciaba una larga y estrecha lengua de sedimentos relativamente ricos en metal. Una segunda lengua se proyectaba, esta vez hacia el Este, a una latitud de 30 grados Sur. Con posterioridad, John E. Lupton y Harmon Craig, de la Institución Scripps, delimitaron una espectacular pluma de helio 3 que se extendía hacia el Oeste desde el eje de la dorsal, a través de buena parte del Pacífico y también a una latitud de 15 grados Sur.

Nadie atinaba a explicar tales observaciones, hasta la reciente publicación de un informe de Joseph L. Reid, de la Scripps. Reid había registrado sistemáticamente la densidad del agua del océano en muchos enclaves. Con los datos recogidos dedujo las fuerzas de cizalla que se generaban entre las diferentes capas del océano. Comprobado, por tanto, que en el interior de los océanos existían "superficies" donde el agua se conservaba inmóvil (que corresponden a límites entre corrientes que circulan en direcciones diferentes), podía calcular las magnitudes y direcciones del flujo respecto a una cualquiera de tales superficies. Reid cartografió, por este procedimiento, el movimiento del agua del Pacífico meridional a una profundidad de 2000 metros, respecto a un "nivel de movimiento cero", que se consideraba situado a 3500 metros. Asombrosamente, las asimetrías encontradas por Boström coincidían con las líneas de flujo deducidas por Reid. Recibía una justificación también la pluma de helio 3. Efectivamente, el helio es un elemento químico no reactivo, y no experimenta, pues, cambio alguno en su arrastre por las corrientes oceánicas. De hecho, "marca" las aguas oceánicas que proceden de una zona con afluencias hidrotermales.

Depósitos de tipo Besshi

Los descubrimientos registrados en los ejes de las dorsales han contribuido en gran medida al conocimiento de los procesos que forman los depósitos minerales en las ofiolitas; en igual proporción han ayudado a la mejor comprensión de un tipo de depósito continental de gran valor económico. Nos referimos a los depósitos de tipo Besshi, nombre otorgado en memoria del que se halló en Japón. Constan de un nivel de sulfuros, muy grueso, localizado entre sedimentos de tipo argilita, roca que procede de una arcilla de grano fino. El mineral no se sitúa, por consiguiente, a techo del basalto. La verdad es que hay que perforar la argilita varios centenares de metros bajo el depósito mineral para llegar al basalto, que por otro lado se encuentra en forma de diques y diques-capa, esto es, como intrusiones de magma, verticales y horizontales, respectivamente.



10. DEPOSITO DE TIPO BESSHI. Se trata de un acúmulo de mineral localizado entre rocas sedimentarias. Se sitúa su origen bajo el océano, en un eje dorsal, próximo a un área emergida que soterraron los derrubios de origen erosivo continental. Los sulfuros ricos en metales se concentran centenares de metros por encima de los diques y diques-capa basálticos: intrusiones verticales y horizontales de magma que generan actividad hidrotermal, a la vez que se enfrían, tras su emplazamiento entre los sedimentos.

Un depósito de tipo Besshi sólo puede formarse cuando el eje de la dorsal se halla próximo a una región emergida que aporta el enorme caudal de sedimentos detríticos de origen erosivo que lo acompañan. Se conoce un ejemplo que hoy está en plena formación. En la cuenca de Guaymas, en el centro del golfo de California, la dorsal del Pacífico oriental penetra bajo la placa continental. Los aportes fluviales procedentes de México han cubierto el eje de la dorsal con sedimentos muy finos, que alcanzan un espesor de varios centenares de metros. En esas circunstancias, el emplazamiento de nueva corteza oceánica adquiere características muy especiales. El magma ascendente no fluye ya al fondo del océano, sino que se inyecta entre los sedimentos detríticos finos, formándose diques y diques-capa. Por tratarse de unos sedimentos bastante permeables al agua, se registra una intensa actividad hidrotermal. De hecho, Lawrence A. Lawver, de la Scripps, pudo localizar el lugar gracias al elevado flujo térmico submarino existente en la zona.

Con posterioridad, Peter F. Lonsdale, de la Scripps, encontró en el centro de la cuenca de Guaymas lentejones de sulfuros que alcanzaban varios metros de altura por centenares de metros de longitud. La fusión de la envoltura de plástico de los testigos extraídos por Lonsdale indicaba temperaturas de 100 grados Celsius a tan sólo 10 metros bajo el fondo del océano. Por otro lado, John Lupton, también de la Scripps, encontró en ese mismo punto elevadas concentraciones de helio 3. Las perforaciones profundas cortaron, efectivamente, diques-capa basálticos.

En enero de 1982 participamos en una expedición dirigida por Lonsdale en esa misma área. Con siete inmersiones, el *Alvin* muestreó aguas surgentes de tres de los lentejones de mineral de mayor tamaño. Las más calientes alcanzaban los 315 grados Celsius; la concentración, muy baja, de magnesio y de sulfato permitía encuadrarlas en la fase final del sistema. Sin embargo, su composición difería radicalmente de la que ofrecían las aguas surgentes no diluidas de los ejes de las dorsales en mar abierto. Diferencia que cabía atribuir a su ascenso a través de los sedimentos detríticos finos de Guaymas. Las aguas surgentes puras, que al abandonar la zona de reacción basáltica deben ser bastante ácidas (al igual que las aguas surgentes en las chimeneas negras situadas a 21 grados de latitud Norte) disolverían las conchas de carbonato cálcico del plancton enterradas en los sedimentos. La disolución del carbonato, que alcaliniza el agua, induce la precipitación de los sulfuros. Al entrar en contacto con el agua oceánica, las aguas surgentes no diluidas de Guaymas son unas cuatro veces más alcalinas que aquélla, y aparecen desprovistas de hierro y de otros elementos formadores de menas.

La diferencia esencial de Guaymas reside en la riqueza de la vida planctó-

nica que muestra aquí el océano. En las primeras, las aguas surgentes viran a alcalinas bajo el fondo del océano, quedando así atrapado el metal contenido en la disolución. ¿Qué ocurre sobre los ejes de las dorsales, en mar abierto, como en el caso ya citado a 21 grados de latitud Norte? Aquí, la alcalinización tiene lugar en el momento de producirse el contacto con el agua del mar. Se forman así las chimeneas negras y se produce la consiguiente dispersión de la mayor parte del contenido metálico de la disolución.

Olor a aceite pesado

En Guaymas, la riqueza del plancton parece haber producido una curiosa modificación del ecosistema vinculado al campo de surgencias. Cuando Lonsdale extrajo los testigos de sedimentos, observó que emanaban un fuerte olor, semejante al del aceite pesado. Al volver, el año pasado, los portamuestras de agua del *Alvin* se quedaban frecuentemente obstruidos por glóbulos de cera. Estos datos tienen fácil explicación. El sedimento fino de Guaymas es rico en carbono planctónico, que produce hidrocarburos por acción del calor desprendido del campo hidrotermal. Este área se distingue, además, por sus grandes colonias de bacterias, que presumiblemente obtienen el alimento de los propios hidrocarburos. Las bacterias, a su vez, parecen competir con auténticos racimos de gusanos tubícolas alrededor de las bocas de las surgencias. Pero no medran bivalvos, mejillones, peces ni pequeños cangrejos, aunque son corrientes ciertos cangrejos cornamentados que parecen depredar a los gusanos. El campo de surgencias de Guaymas, en respuesta a su actividad hidrotermal específica, difiere en todos los sentidos de cuanto se ha observado en los ejes de dorsales situadas en mar abierto, lo que hace presagiar una gran diversidad en futuros descubrimientos.

Las últimas dos décadas han sido testigo de una verdadera explosión en las ciencias planetarias y de la Tierra. En nuestro planeta, la teoría de la tectónica de placas ha unificado la geología, hasta hace poco una ciencia fragmentada y descriptiva. El descubrimiento de la actividad hidrotermal en los ejes de las dorsales submarinas demuestra, además, que lo novedoso y lo original no son aspectos privativos de las misiones espaciales a otros planetas. Se dice que la Tierra es el planeta del agua. Podemos afirmar ahora que la química del agua y de los sedimentos oceánicos lleva, a su vez, la impronta de los procesos volcánicos.

Una ventana abierta al cerebro dormido

El sueño REM, fase en la que se producen las ensoñaciones, suele ir acompañado de parálisis. La inhibición del bloqueo muscular en animales de experimentación permite investigar esos episodios

Adrian R. Morrison

Suele considerarse el sueño un estado de completa relajación e inactividad. Sin embargo, al observar la actividad eléctrica desencadenada en el cerebro de un ser humano o un animal de experimentación dormidos, descubrimos que el sueño es un estado complejo y en absoluto quieto. Es más, la actividad eléctrica cerebral en la fase que se producen las ensoñaciones –el sueño REM (del inglés "*r*apid *e*ye *m*ovement", movimiento ocular rápido)– se asemeja al estado de vigilia mucho más que cualquier otra fase del sueño. En el REM, los cambios transitorios del voltaje indican que el cerebro está despierto, y ello pese a que el contacto sensorial entre el animal y su entorno sufre una reducción considerable. Sólo impide que el despertar neurológico del cerebro en el sueño REM se traduzca en movimiento físico vigoroso la ausencia de actividad de los músculos del animal; lo que se manifiesta en una parálisis que persiste hasta finalizar el episodio REM.

¿Qué ocurriría si, eliminando tal parálisis, el cerebro activara las fibras musculares? Los trabajos efectuados en mi laboratorio de la Universidad de Pennsylvania, y en otros centros, aportan datos para elaborar la respuesta. La parte más primitiva es el tronco cerebral, situado entre la médula espinal y el resto del cerebro. Si provocamos pequeñas lesiones en el tronco cerebral del gato, el animal presenta las características eléctricas del sueño REM, pero se mueve con vigor. Las lesiones se practican en la región del tronco cerebral denominada protuberancia; la posición exacta de la lesión determina qué músculos quedan liberados de la parálisis.

La carencia de tono muscular durante el sueño REM parece, por tanto, directamente atribuible a la protuberancia. Los resultados de otros experimentos sugieren que las neuronas (células nerviosas) de la protuberancia afectan también a la locomoción, si bien indirectamente. En los episodios REM, una región de la protuberancia parece interferir con la acción de un "centro locomotor" que se extiende desde la protuberancia hasta un punto inferior del tronco cerebral. Esa interferencia impide que el centro locomotor active las redes nerviosas de la médula espinal responsables del movimiento recíproco (alternante) de los miembros. La actividad de la protuberancia conduce, por tanto, de dos maneras distintas a la inhibición del movimiento. Mi propio trabajo sugiere que ese eslabón que engarza el aumento de la actividad cerebral con la aparición de parálisis muscular actúa durante la vigilia y que quizá sea la causa de ciertos trastornos del sueño.

La transición de un estado de vigilia a otro de sueño comporta, en el animal, numerosos cambios fisiológicos y posturales; el efecto conjunto de esas alteraciones se traduce en un profundo cambio de estado. Gran parte de mi trabajo lo he realizado con gatos, y los cambios aquí descritos se referirán específicamente a ellos, pero la mayoría pueden aplicarse a otros mamíferos. Cuando un gato cae dormido, adopta la postura enroscada y relajada que todos hemos visto. Se le cierran los párpados y la membrana nictitante (el "tercer párpado") cubre parte del ojo, por debajo de aquéllos. Se reduce el número de fibras musculares en contracción y, por tanto, se rebaja el tono muscular. En la fase inicial del sueño desaparecen los movimientos oculares apreciables en el animal despierto. Se produce asimismo un gradual descenso de la temperatura cerebral, generalmente en una fracción de grado Celsius. Subyacen a esos cambios las alteraciones de la actividad neurológica que se desencadenan al iniciarse el sueño. La información más completa sobre la actividad eléctrica del cerebro nos la proporciona el electroencefalograma (EEG). Se trata de un registro de los continuos cambios del potencial eléctrico, medido entre electrodos fijados al cráneo. Los rápidos cambios del potencial representan la suma de los cambios de voltaje que se dan en la membrana externa de un buen número de neuronas cerebrales.

En el EEG de un gato despierto se aprecian ondas con una amplitud bastante baja y una frecuencia bastante alta. Cuando el gato se enrosca y se dispone a dormir se observa un cambio notable. Las ondas de baja amplitud del estado de vigilia se reemplazan gradualmente por ondas de gran amplitud y baja frecuencia. La fase inicial del sueño ligero suele denominarse sueño de ondas lentas, precisamente por la existencia de ese patrón de baja frecuencia en el EEG. A medida que entra en esa fase, se descubren cambios de la actividad eléctrica en áreas específicas de los cerebros superior e inferior; tales cambios se registran mediante la inserción de electrodos en las estructuras cerebrales adecuadas.

Durante muchos años se creyó que ese patrón de ondas lentas persistía en todo el sueño. Sin embargo, en 1953, Eugene Aserinsky y Nathaniel Kleitman, de la Universidad de Chicago, observaron en humanos intervalos de sueño en los que el registro EEG retornaba al patrón de alta frecuencia y baja amplitud de la vigilia. Tras el cambio, advertíase que los individuos movían con rapidez los ojos en distintas direc-

ciones; de ahí que se denominara "movimiento ocular rápido" esa fase del sueño. Sabemos ahora que se presentan regularmente fases de sueño REM, alternándose con períodos, más largos, de sueño de ondas lentas. La localización temporal y la duración de esos episodios varían según las especies. Las fases REM ocurren, en los humanos, aproximadamente cada 90 minutos, y cerca de cada 25 minutos en los gatos. Los episodios duran varios minutos cada uno.

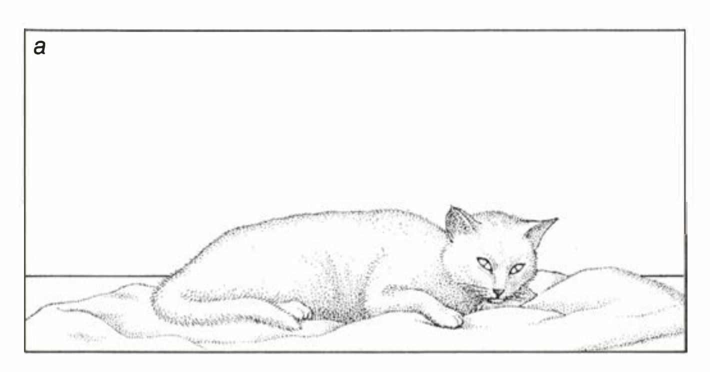
Aserinsky y Kleitman demostraron que el REM va asociado a las ensoñaciones, vívidas e intensas; pudieron comprobarlo despertando a sujetos experimentales inmediatamente después de que acabara una de esas fases. La identificación del momento en que se produce la ensoñación constituyó uno de los puntos que suscitó mayor interés tras el descubrimiento de la fase REM. Trabajos posteriores han demostrado, sin embargo, que los aspectos fisiológicos del sueño REM no encierran un atractivo menor.

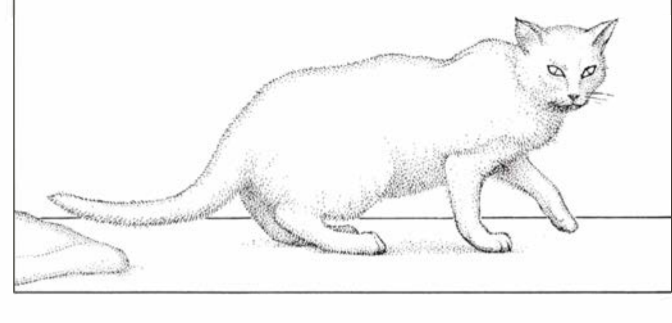
Además de la práctica concordancia de los patrones EEG en el sueño REM y en la vigilia, el ritmo de actividad de las neuronas de la mayoría de las subunidades del cerebro son bastante similares en ambos estados. Uno de los parecidos más estrechos se da en el hipocampo, que es, filogenéticamente, la porción más antigua de la corteza cerebral. A lo largo de gran parte del sueño REM y de la vigilia se presenta un patrón regular de ondas en el hipocampo, a un ritmo aproximado de siete por segundo. Ese patrón, el "ritmo theta", difiere absolutamente de la actividad del hipocampo en el sueño de ondas lentas, menos regular y con ondas picudas. También es parecida la temperatura del cerebro: después de descender una fracción de grado en el sueño de ondas lentas, la temperatura cerebral aumenta súbitamente, hasta alcanzar el nivel de la vigilia, durante un episodio de sueño REM.

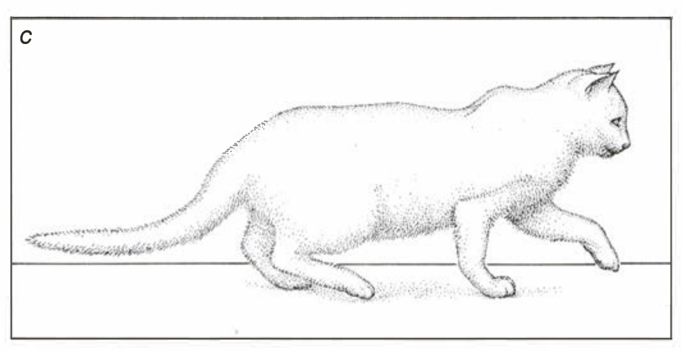
Muchas características del sueño REM guardan, pues, un señalado paralelismo con la vigilia y se apartan completamente de las características del sueño de ondas lentas. Es de subrayar que todos los signos mencionados antes corresponden a la activación de la formación reticular, o sistema activador reticular, situada en el centro del tronco cerebral. En 1949, Giuseppe Moruzzi y Horace W. Magoun, que trabajaban juntos en la Universidad del Noroeste, demostraron que la formación reticular era responsable del despertar. En los episodios REM, por tanto, el cerebro aparece fuertemente activado. Dada la yuxtaposición de descanso y actividad, el sueño REM se denomina, también, sueño paradójico.

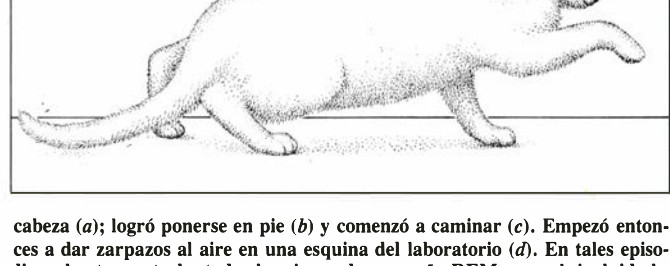
La íntima conexión del REM con la vigilia se corroboró recientemente en un inesperado descubrimiento logrado en mi laboratorio. Uno de mis alumnos, Robert Bowker, trabajaba sobre las ondas eléctricas que aparecen "espontáneamente" (es decir, sin estimulación externa conocida) en el sueño REM. Esas ondas, breves y de gran amplitud, se denominan picos PGO, por referencia a las estructuras cerebrales donde se han estudiado: la protuberancia (de donde se supone que proceden) y dos partes del sistema visual, el cuerpo geniculado lateral y el córtex visual (la región más occipital del córtex visual).

Se creía que los picos PGO aparecían únicamente durante el sueño REM, y que en escasas ocasiones se producían en el sueño de ondas lentas. Bowker golpeó sin querer el electroencefalógrafo mientras se grababa el sueño de ondas lentas de un gato. Casi inmediatamente apareció un pico PGO en el registro. El estudio subsiguiente demostró que podían provocarse a voluntad picos PGO, tanto en el sueño REM como en el sueño de ondas lentas, mediante estímulos sonoros o táctiles. Se interpretaban, por tanto, los picos

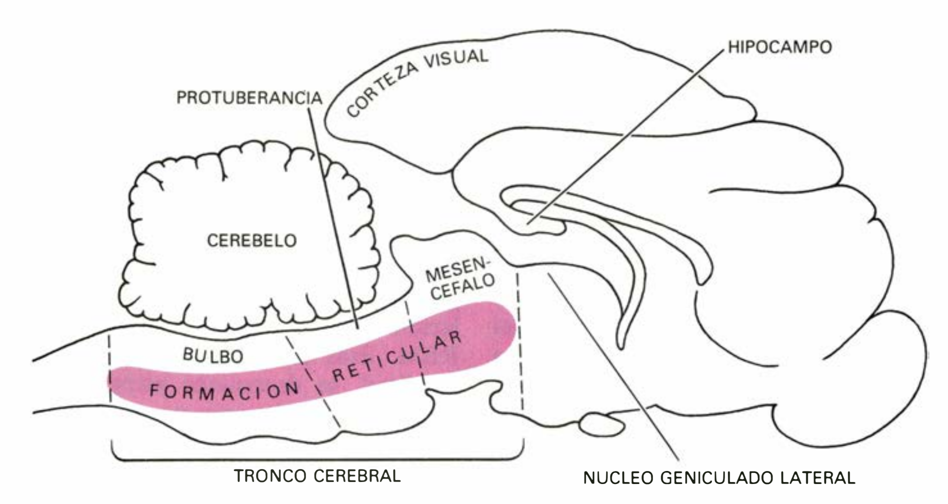








1. SUEÑO REM SIN PARALISIS MUSCULAR, mostrado a través de dibujos basados en las películas rodadas en el laboratorio del autor, en la Universidad de Pennsylvania. La parálisis que normalmente acompaña al sueño REM se eliminó practicando una pequeña lesión en el tronco cerebral del gato [*véase la figura 7*]; el grado con que se reducía la parálisis variaba con la lesión. En el episodio de sueño REM que se reproduce aquí, el gato erguió la cabeza (*a*); logró ponerse en pie (*b*) y comenzó a caminar (*c*). Empezó entonces a dar zarpazos al aire en una esquina del laboratorio (*d*). En tales episodios, el gato mostraba todos los signos de un sueño REM normal, incluida la extensión de la membrana nictitante ("el tercer párpado") que le cubría la mayor parte del ojo. La ligera contracción que exhibe el animal en los cuartos traseros evidencia que no se ha liberado completamente el tono muscular.



2. ESTRUCTURAS CEREBRALES que desempeñan algún papel en el sueño, representadas en una sección esquemática del cerebro de gato. A la derecha, la parte frontal del cerebro. Las lesiones que eliminan la parálisis en el REM afectan a la protuberancia, estructura ubicada en el tronco cerebral.

PGO, considerados hasta entonces fenómenos eléctricos espontáneos limitados a la fase REM, como respuestas multisensoriales de alerta evocables en distintos estados cerebrales.

Tal conclusión indujo a Bowker a reexaminar las ondas denominadas PMO (potenciales de movimientos oculares), que aparecen en el estado de vigilia. El registro eléctrico de tales ondas es idéntico a los picos PGO, y se observan en las mismas estructuras cerebrales. Se atribuía la presencia de PMO al nivel de iluminación del entorno del gato, ya que esos potenciales no se habían observado en gatos despiertos encerrados en una estancia oscura. Se admitió que la interrelación entre los PMO y la iluminación del entorno constituía una razón significativa de discrepancia con respecto a los picos PGO.

Razonó Bowker que un gato encerrado en una habitación de registro a oscuras debía aburrirse bastante, y se le ocurrió despertar el interés del animal enviándole el aroma de un atún. No tardaron en producirse ondas PMO, que eran idénticas a los picos PGO. Igual resultado se logró emitiendo breves ráfagas de un sonido intenso; de lo que se deduce que tanto los picos PGO como las ondas PMO son formas particulares de una respuesta general de alerta. La respuesta se evoca, por estimulación, durante la vigilia, en el sueño de ondas lentas o en el REM. En éste, según parece, también puede darse espontáneamente la activación de esas mismas neuronas. Durante la fase REM el cerebro se activa hasta alcanzar un estado similar al de vigilia, aunque el contacto sensorial con el entorno sea muy pobre.

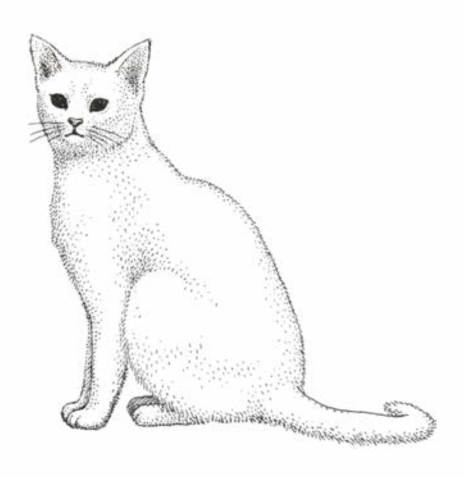
¿Qué protege de la lesión a un animal en sueño REM, cuyo sistema nervioso central está activo y alerta, aunque desconectado del entorno en gran parte? Precisamente en esas fases se inhibe la acción de las neuronas motoras de la médula espinal, que provocan la contracción muscular voluntaria. Los músculos se muestran atónicos, esto es, sin tono, y el animal queda paralizado. (Debe señalarse que parálisis y atonía no son sinónimos; la parálisis puede resultar de diversas afecciones. En todo caso, si los músculos pierden el tono, el animal queda paralizado.)

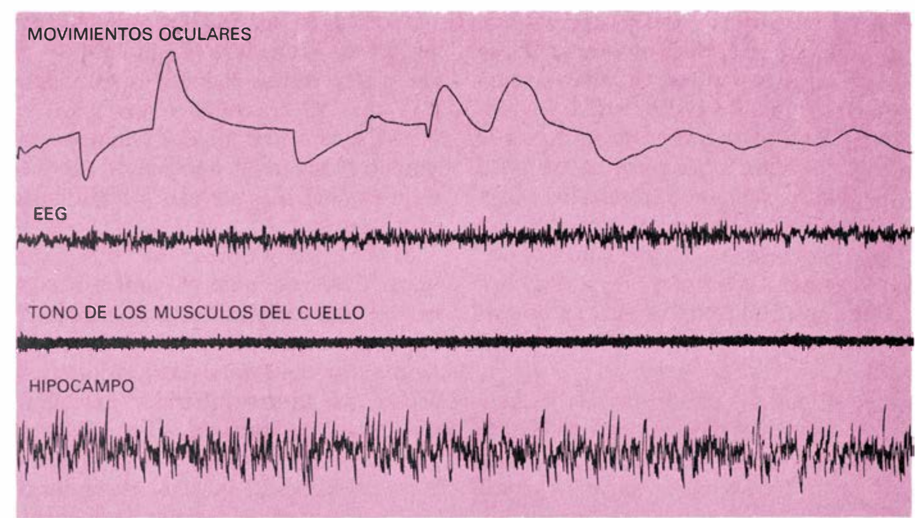
El grado de actividad de los músculos esqueléticos (o voluntarios) suele registrarse por medio de electrodos insertos en el cuello. Los músculos del cuello se cuentan entre los "antigravitatorios", responsables del mantenimiento de la postura erecta. Como se ha visto, cuando el gato entra en sueño ligero, baja el tono de sus músculos, pues se reduce el número de fibras activas en cada músculo. Cuando el animal entra en una fase de sueño REM, el registro del tono muscular es plano, lo que indica una falta absoluta de tono. Hay, sin embargo, descargas ocasionales de actividad muscular, que producen sacudidas en varias partes del cuerpo. En mi laboratorio hemos trabajado con gatos a los que se devolvía el tono durante el sueño REM por medio de lesiones practicadas en la protuberancia; destruíamos un pequeño volumen de tejido con un alambre caliente que introducíamos en el cerebro según unas coordenadas predeterminadas. Tras provocar la lesión, se observaba a veces una notable actividad, que arrancaba transcurrido un intervalo de ondas lentas, cuando el gato debía entrar en una fase REM con atonía. El animal levantaba la cabeza, estiraba el cuerpo, movía alternativamente los miembros e intentaba levantarse; algunos lo conseguían, otros incluso andaban. Los gatos que sufrían ciertas lesiones mostraban comportamientos habituales de la vigilia: exhibían gestos típicos de orientación hacia la presa, la buscaban y atacaban. Tales conductas raramente se dirigían hacia objetos del entorno.

Como he señalado, el tipo de actividad muscular que desarrollaba el gato en esos extraordinarios episodios dependía del tamaño y la posición de las lesiones cerebrales. Al examinar al microscopio óptico secciones cerebrales de un gato que sólo lograba sostenerse en sus patas delanteras se observaron lesiones pequeñas y simétricas en la parte más dorsal y superior de la protuberancia. En un gato que se apoyaba sobre las cuatro patas y caminaba se apreciaron lesiones mayores en una posición inferior y más ventral. Un gato que exhibía una conducta agresiva, golpeando repetidamente el suelo, presentaba lesiones que se extendían hacia delante y penetraban en el mesencéfalo. La significación de esas ubicaciones se aclarará cuando examinemos las vías nerviosas implicadas en el sueño REM.

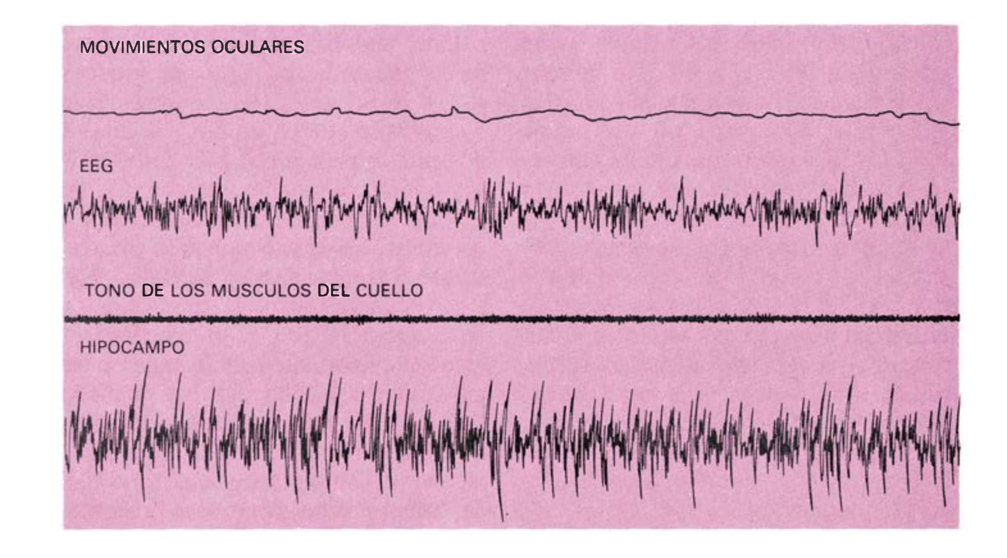
Varios detalles sugieren que esos movimientos constituyen, en efecto, un sueño REM sin atonía. El fenómeno fue identificado en primera instancia por Michel Jouvet y François Delorme, de la Universidad de Lyon. Como en el sueño REM ordinario, el EEG de los gatos que efectúan alguna actividad mientras duermen se parece al de la vigilia. La membrana nictitante cubre parcialmente el ojo y las pupilas se cierran dejando sólo un resquicio. En los registros obtenidos del hipocampo se aprecia el ritmo theta. Mi alumna Joan Hendricks sujetó gatos dormidos con un aparejo almohadillado y observó que los ejemplares lesionados efectuaban movimientos rápidos de los ojos, bigotes y dedos, propios del sueño REM; de haber estado libres, lo normal hubiera sido que iniciaran el repertorio de movimientos complejos. Comprobó que la temperatura cerebral se elevaba, como ocurre en la fase REM normal.

Hendricks demostró también que en el sueño REM sin atonía se daba asimismo una característica particularmente intrigante del sueño REM ordinario. Pier Luigi Parmiggiani y sus colegas, de la Universidad de Bolonia, habían demostrado que, durante el

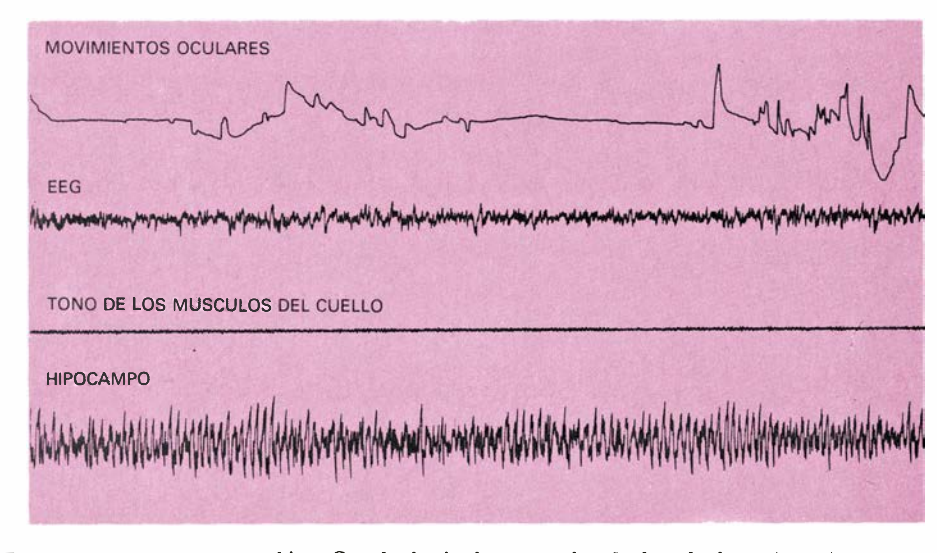












3. TRES ESTADOS (vigilia, sueño lento y sueño REM) pueden detectarse en los gatos por los cambios de postura, tono muscular y actividad cerebral. En vigilia (*arriba*) los ojos y la cabeza se mueven al responder el animal a los estímulos visuales y auditivos. El electroencefalograma, que registra los cambios de potencial entre electrodos fijados al cráneo, muestra un patrón de baja amplitud y ondas de alta frecuencia. Los músculos esqueléticos (voluntarios) presentan un nivel acusado de tono, que se mide habitualmente en el cuello. El hipocampo, un área primitiva de la corteza cerebral, muestra un patrón regular, llamado ritmo theta, cuando el gato presta atención a algún objeto. Cuando el animal entra en el sueño de ondas lentas (*centro*), se enrosca; cesan los movimientos oculares; el EEG registra un patrón de baja frecuencia y gran amplitud; el tono muscular desciende y el hipocampo muestra un ritmo irregular. Cuando el animal entra en el período de sueño REM (*abajo*) la postura enroscada se relaja ligeramente. Reaparecen los movimientos oculares rápidos, el patrón electroencefalográfico de alta frecuencia y el ritmo theta. El tono de los músculos esqueléticos, sin embargo, desaparece completamente, situación que se denomina de atonía. La coincidencia de actividad cerebral y atonía indujo a denominar sueño paradójico a la fase REM.

REM, los gatos perdían la capacidad de regular la temperatura corporal. En esa fase, los gatos parecen animales de sangre fría, como los peces, anfibios y reptiles. Hendricks comprobó que, en el REM sin atonía, los gatos no tiritaban ni erizaban el pelo en respuesta al frío, aunque los mismos ejemplares exhibían tales respuestas si estaban despiertos. Las lesiones de la protuberancia no restauran, por tanto, la capacidad de respuesta al frío.

Gran parte de mi trabajo lo he dedicado a esclarecer por qué las pequeñas lesiones de la protuberancia provocan la desaparición de la atonía del sueño REM y desencadenan movimientos elaborados muy similares a los del animal despierto. Nos hemos centrado en el estudio de la protuberancia porque, en 1962, al poco de comenzar su investigación sobre el sueño, Jouvet demostró que los gatos que carecían de la parte de cerebro que queda por delante de la protuberancia mostraban períodos de atonía y movimientos oculares rápidos idénticos a los característicos del sueño REM. En el curso de mis trabajos he llegado a la conclusión de que existen al menos dos sistemas nucleares separados que actúan en el sueño REM sin atonía. Uno es responsable de la liberación del tono muscular, el otro, de la liberación de la conducción motora, un impulso generalizado hacia la locomoción.

El primer sistema parece ser el más simple. Se sabe que en el sueño REM actúa un centro inhibidor del bulbo raquídeo. (La acción de las neuronas puede ser excitadora o inhibidora. Las áreas inhibidoras reducen la excitabilidad de las neuronas diana.) En el REM, el centro inhibidor del bulbo raquídeo bloquea las neuronas de la médula espinal que activan los músculos esqueléticos.

Si el centro inhibidor del bulbo estuviera, a su vez, bajo el control excitador de una red de células nerviosas de la protuberancia, la lesión de ésta destruiría las conexiones excitadoras; la acción del centro inhibidor del bulbo quedaría interrumpida y los músculos esqueléticos retendrían el tono durante el sueño REM. La posición de las lesiones determinaría el alcance de la liberación del tono y, por tanto, qué musculos podría emplear el gato.

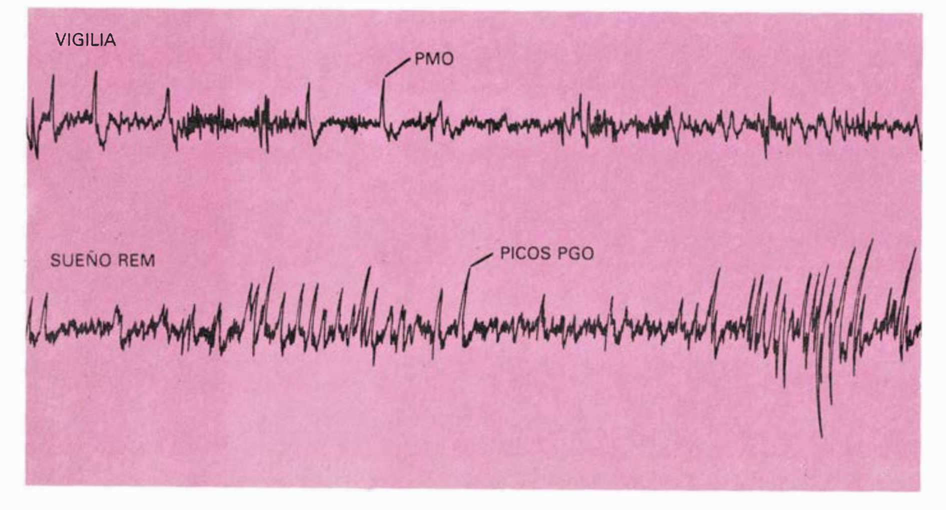
Esta hipótesis ha recibido el apoyo de los resultados logrados por Jouvet y sus colegas en Lyon, quienes han demostrado que en un área situada en posición dorsal, junto a la zona de la protuberancia donde provocamos las lesiones, se origina un tracto de fibras nerviosas. El tracto conduce directamente al área inhibidora del bulbo. Hemos provocado lesiones en el origen y en el curso del tracto, lo que produce gatos que mantienen cierto tono en los músculos del cuello durante el sueño REM, si bien sólo pueden levantar ligeramente la cabeza. Ello sugiere que para liberar completamente el tono han de dañarse otras neuronas, probablemente de la formación reticular.

Los datos experimentales parecen confirmar, por tanto, que bajo la atonía del sueño REM subyace una conexión directa entre la protuberancia, el centro inhibidor del bulbo y los músculos esqueléticos. Puesto que, como se sabe, el sueño REM está asociado en los humanos con intensas ensoñaciones, sería tentador concluir que, rompiendo esa conexión, asistiríamos a la representación de sus sueños por parte de un animal. Además de las dificultades inherentes a la atribución de estados mentales complejos a otras especies, existen buenas razones para suponer que no se acaban ahí los procesos desencadenados durante un episodio de sueño REM sin atonía.

Las funciones del cerebro las desempeñan estructuras anatómicas estrechamente interconectadas. Sería poco razonable suponer que la lesión de un área central del tronco cerebral comportara sólo la destrucción de estructuras inhibidoras de las neuronas motoras. Tenemos datos que prueban que las lesiones de la protuberancia afectan a otros sistemas inhibidores del bulbo raquídeo.

Las pruebas de mayor peso conciernen a la actividad locomotora de gatos lesionados y despiertos. No muestran éstos un aumento anormal del tono muscular durante el sueño de ondas lentas o durante la vigilia. El único efecto sobre el tono se registra durante el sueño REM. No obstante, los gatos parecen más activos que antes de sufrir las lesiones. Cuando se les deja sueltos por el laboratorio se muestran molestos e inquietos, corren de un lado para otro, investigando objetos sin interés.

Para confirmar la presunta existencia de una actividad motora inusual realizamos algunos experimentos. Efectuamos los ensayos con gatos, antes y después de provocarles las lesiones. Los introducíamos en una habitación que tenía el suelo escaqueado. El número de escaques que recorría el gato en un período de 30 minutos servía de medida de la actividad motora. Todos los gatos que tenían sueño REM sin atonía mostraban también un aumento de actividad tras la lesión. Los crecimientos se situaban entre el 30 y el 261 por ciento, y todos eran estadísticamente significativos. Tales datos me llevaron a concluir que las lesiones de la protuberancia afectaban a la fuente de la conducción motora generalizada. Se trata, probablemente, de un sistema anatómico diferente del que afecta al tono muscular, puesto que, cuando los animales estaban despiertos, su tono muscular

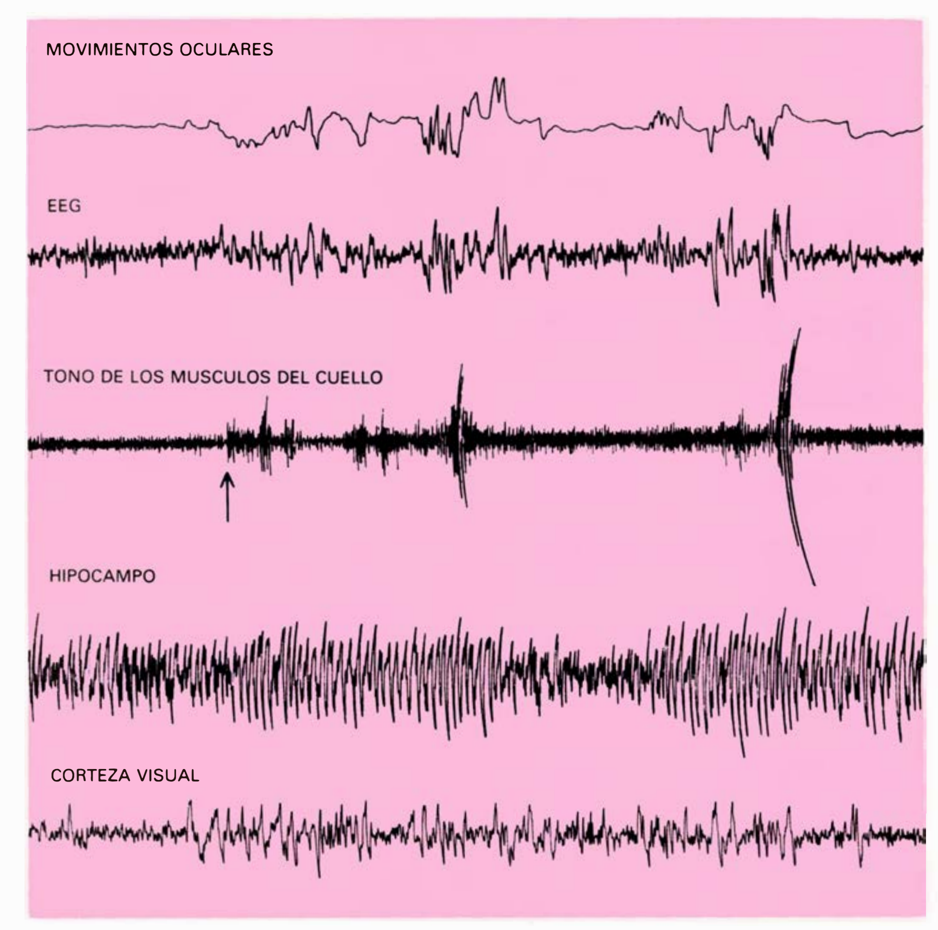


4. ACTIVACION DEL CEREBRO durante la fase REM. Pudo demostrarse recientemente en el laboratorio del autor. En el estado de vigilia aparecen, en el área visual de la corteza cerebral, unas ondas llamadas potenciales de movimientos oculares (PMO) (*arriba*), que se consideraban dependientes sólo del nivel de iluminación. Robert Bowker, uno de mis alumnos, demostró que podían evocarse por medio de una gran variedad de estímulos, incluso en absoluta oscuridad: se trata de respuestas multisensoriales de alerta. Las ondas son idénticas a las conocidas como picos PGO, que aparecen espontáneamente en las mismas áreas cerebrales durante el sueño REM (*abajo*). En el sueño REM, el cerebro se mantiene alerta. El estado de alerta, sin embargo, responde a la actividad cerebral, y no a la información sensorial.

no manifestaba influencia alguna de las lesiones.

Nuestros trabajos nos permiten esbozar ya un cuadro de cómo podría funcionar el segundo sistema de inhibición durante el sueño. Mi hipótesis de la actuación del sistema se apoya en las investigaciones sobre el control nervioso de la locomoción. Se ha demostrado que la región del tronco cerebral por la que nos interesamos puede regular la locomoción sin que intervengan centros cerebrales superiores. En tal regulación se hallan implicadas tres estructuras. Un "generador de locomoción" de la médula, compuesto por neuronas que controlan y coordinan el movimiento recíproco de los miembros en la marcha y la carrera, y que, a su vez, está bajo el control excitador de una segunda estructura: "un centro locomotor" del tronco cerebral. Si se estimula el centro locomotor, éste incita al generador locomotor. La tercera estructura es una región de control de la protuberancia, conectada al centro locomotor por un enlace inhibidor. Al activarse esta región, se suprime la actividad del centro locomotor. Así, en el sueño REM, la protuberancia suprimiría indirectamente la actividad motora y el tono muscular; la lesión de la protuberancia liberaría los músculos locomotores.

La existencia del centro locomotor del tronco cerebral y del generador de locomoción espinal ha sido postulada por numerosos autores, pero se desconoce aún su organización anatómica. La primera prueba de su existencia la recabaron M. L. Shik, F. V. Severin y G. N. Orlovskii, de la Universidad estatal de Moscú; sus resultados los han corroborado otros autores. Shik y sus colegas han demostrado que la aplicación de una corriente eléctrica a un sistema neuronal que parte del núcleo cuneiforme, ubicado en la región caudal del mesencéfalo, puede provocar, en gatos descerebrados, la reacción de caminar, y aún trotar o galopar, sobre un cilindro rodante. Sospecho que es ese centro el que la lesión de la protuberancia libera de su inhibición.

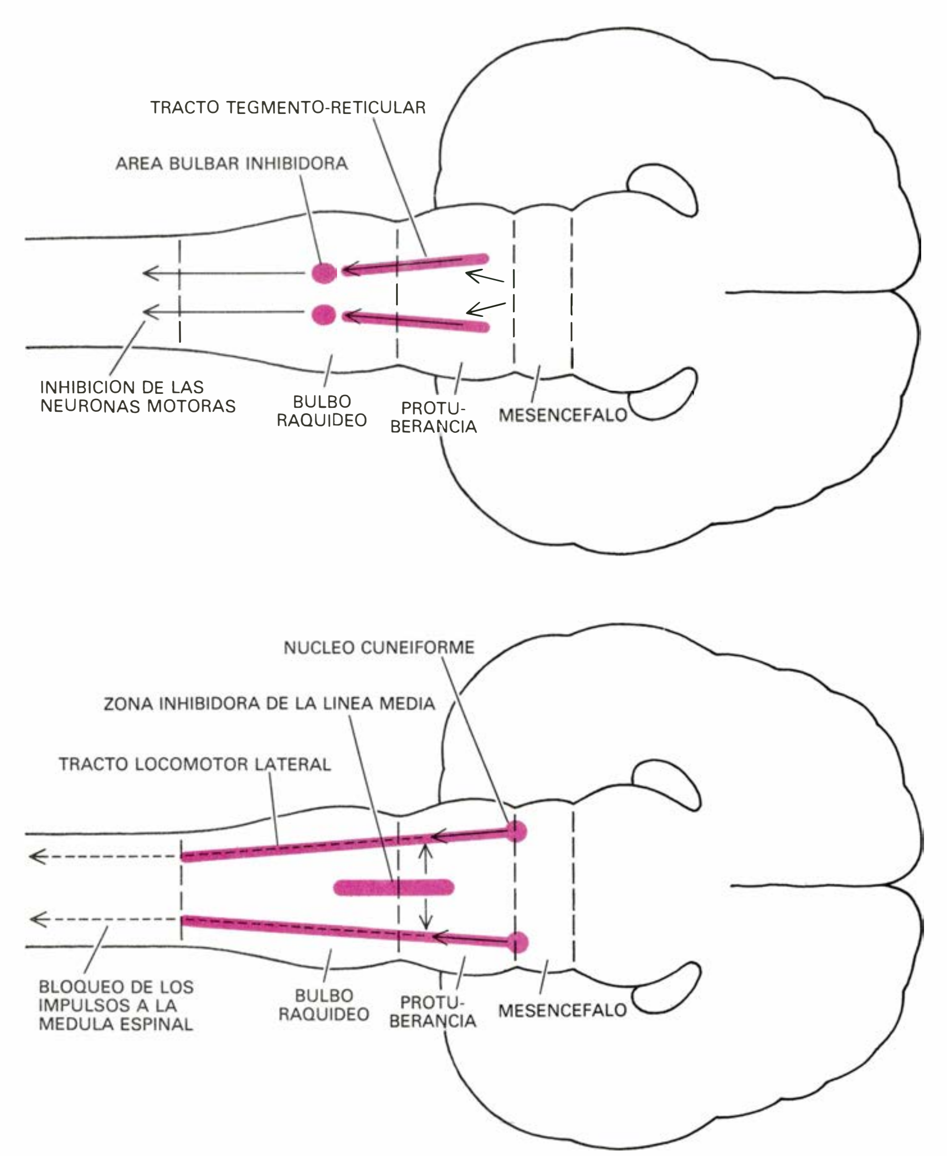


5. EPISODIOS DE MOVIMIENTO DURANTE EL SUEÑO en gatos lesionados. Aparecen cuando todos los signos, salvo el tono muscular, son similares a los del sueño REM. Las similitudes han llevado al autor a concluir que los episodios de movimiento corresponden, en efecto, a fases REM sin atonía. En el sueño REM normal se aprecian movimientos rápidos oculares, un patrón EEG de baja amplitud, ritmo theta y picos PGO. Todos aparecen en el registro aquí mostrado, realizado durante los episodios con movimiento que siguen a la lesión provocada en la protuberancia. La diferencia más notable estriba en que, en el sueño REM sin atonía, los músculos esqueléticos del gato recuperan el tono, permitiendo que el animal se desplace. La flecha indica el comienzo de un episodio de movimiento durante el sueño REM.

Los recientes trabajos de Shigemi Mori y sus colaboradores, de la Facultad de Medicina de Ashikawa, apoyan la existencia del centro locomotor y sugieren cuál podría ser la conexión del centro con la protuberancia. Estos autores han observado que los tractos capaces de inducir el movimiento en gatos descerebrados parten del mesencéfalo hacia el exterior del tronco cerebral, en forma de un par de tractos delgados que atraviesan, por los laterales, la protuberancia y el bulbo. Han demostrado, además, que en la parte central del tronco cerebral existen áreas capaces de facilitar o inhibir el funcionamiento locomotor.

Que el centro locomotor se excite o inhiba dependerá de la estimulación de un área concreta. La estimulación eléctrica de la zona facilitadora más ventral induce movimientos en las patas traseras del animal sin la estimulación directa del propio centro locomotor. La estimulación más dorsal reduce el tono muscular e inhibe la locomoción. Las zonas facilitadoras e inhibidoras descubiertas por Mori y sus colegas están cerca de la línea media del tronco cerebral y próximas a las regiones de la protuberancia donde hemos provocado las lesiones en nuestros experimentos. Se esclarecen así las conexiones entre la protuberancia y el supuesto centro locomotor.

Hay que considerar tales hallazgos a la hora de interpretar los efectos de las lesiones causadas en la protuberancia. Por ejemplo, en el gato mencionado antes, que era capaz de tenerse en pie y caminar, la lesión interrumpía todo el efecto de la zona inhibidora de la línea media sobre el centro locomotor. En el gato capaz de apoyarse sólo sobre las patas delanteras, la lesión no lograba bloquear por entero el efecto inhibidor de la línea media. Mi colaboradora Graziella Mann y yo mismo hemos encontrado recientemente que la gravedad del daño de la zona inhibidora del gato capaz de levantarse sólo sobre las patas delanteras le permite caminar sobre sus cuatro extremidades durante el sueño REM sin atonía. Mas las supuestas conexiones anatómicas en la protuberancia, y entre el centro locomotor y el generador espinal, continúan moviéndose en el terreno de la hipótesis. Debe aún profundizarse en el estudio de esta cuestión. En los gatos que muestran una conducta agresiva se pre-



6. CONEXIONES RETICULARES DE LAS NEURONAS que parecen responsables de la inhibición del tono muscular y de la conducción motora en el sueño REM. Las acciones de las neuronas sobre sus objetivos pueden ser inhibidoras o excitadoras. Se sabe que una zona del bulbo raquídeo que inhibe las neuronas motoras desempeña un papel importante en la reducción del tono muscular (*esquema superior*). El tracto tegmento-reticular conecta la protuberancia al centro inhibidor. En el sueño REM, la protuberancia se activa, excitando el centro bulbar por esa y otras vías. El centro bulbar inhibe las neuronas motoras y provoca la atonía. Un tracto locomotor lateral desempeña una parte importante en la reducción de la conducción motora (*abajo*). El tracto discurre hacia abajo y sale del tronco cerebral. Está conectado a las estructuras de la médula espinal y puede inducir movimientos de las patas en gatos a los que se les ha extirpado la porción de cerebro anterior al mesencéfalo. En el centro del tronco cerebral hay neuronas que excitan o inhiben el tracto locomotor. Se disponen en zonas inhibidoras y facilitadoras, en un plano vertical próximo a la línea media del cerebro. En el sueño REM, la protuberancia estimula la zona inhibidora, bloqueando el tracto locomotor y anulando la conducción motora. La lesión de aquella puede así eliminar el control inhibidor en ambos sistemas, liberándose el tono muscular y la conducción motora.

sentan otras complicaciones. Tienen dañadas las vías neuronales que descienden del cerebro, que, como se sabe, afectan al control de la agresión en la vigilia.

No hay que esperar, sin embargo, que el total esclarecimiento de las conexiones anatómicas específicas conlleve la comprensión del proceso de control de los músculos durante el sueño REM. Quizás operen mecanismos más sutiles que los estudiados hasta ahora. René Drucker-Colin, de la Universidad de México, y sus alumnos Gloria Arankowsky y Raúl Aguilar, como también nuestro grupo, han demostrado recientemente que las descargas neuronales de alta frecuencia que se producen en muchas regiones del cerebro durante el sueño REM son necesarias para la liberación de la actividad durante el REM sin atonía.

La administración de cloranfenicol, un antibiótico común, a dosis comparables a las empleadas en el tratamiento de infecciones bacterianas, provoca la reaparición de la atonía en gatos con lesiones en la protuberancia que, de otro modo, mostrarían episodios de movimiento en el sueño REM. Los gatos se comportan con normalidad cuando están despiertos. Drucker-Colin había demostrado con anterioridad que el cloranfenicol reducía el ritmo de disparo de las neuronas de la formación reticular. No se conoce el mecanismo que induce la reducción de la actividad neuronal. Se sabe, eso sí, que el cloranfenicol inhibe la síntesis proteica: quizás ese efecto explique la reducción del ritmo de disparo de las neuronas. Cualquiera que sea el mecanismo de acción de la droga, tras su administración los gatos lesionados no parecen gozar de impulso motor suficiente para generar la compleja actividad que, de otra forma, exhibiría el sueño REM.

Los trabajos efectuados hasta ahora sobre el sueño REM sin atonía señalan la existencia de un eslabón entre el estado de despertar de la fase REM y la reducción de la actividad motora coordinada. Obviamente, el organismo se beneficia de un mecanismo que impide el movimiento cuando el cerebro es muy activo, pero poco receptivo a la estimulación externa. No se precisan, sin embargo, argumentos teleológicos para explicar la existencia de conexiones entre la reacción de despertar del cerebro y la inhibición motora. Tales conexiones podrían existir aún en vigilia.

Al enfrentarse a un estímulo inesperado, las personas y los animales suelen vacilar antes de actuar. La mayoría habrá experimentado un momento de duda o incluso una sensación de debilidad en las rodillas al pasar junto a ellos un automóvil a gran velocidad. Dicha inhibición motora de la marcha es generalmente efímera. Sin embargo, sugiere que aún en la vigilia hay una conexión entre las respuestas momentáneas de alerta y la reducción de la actividad motora.

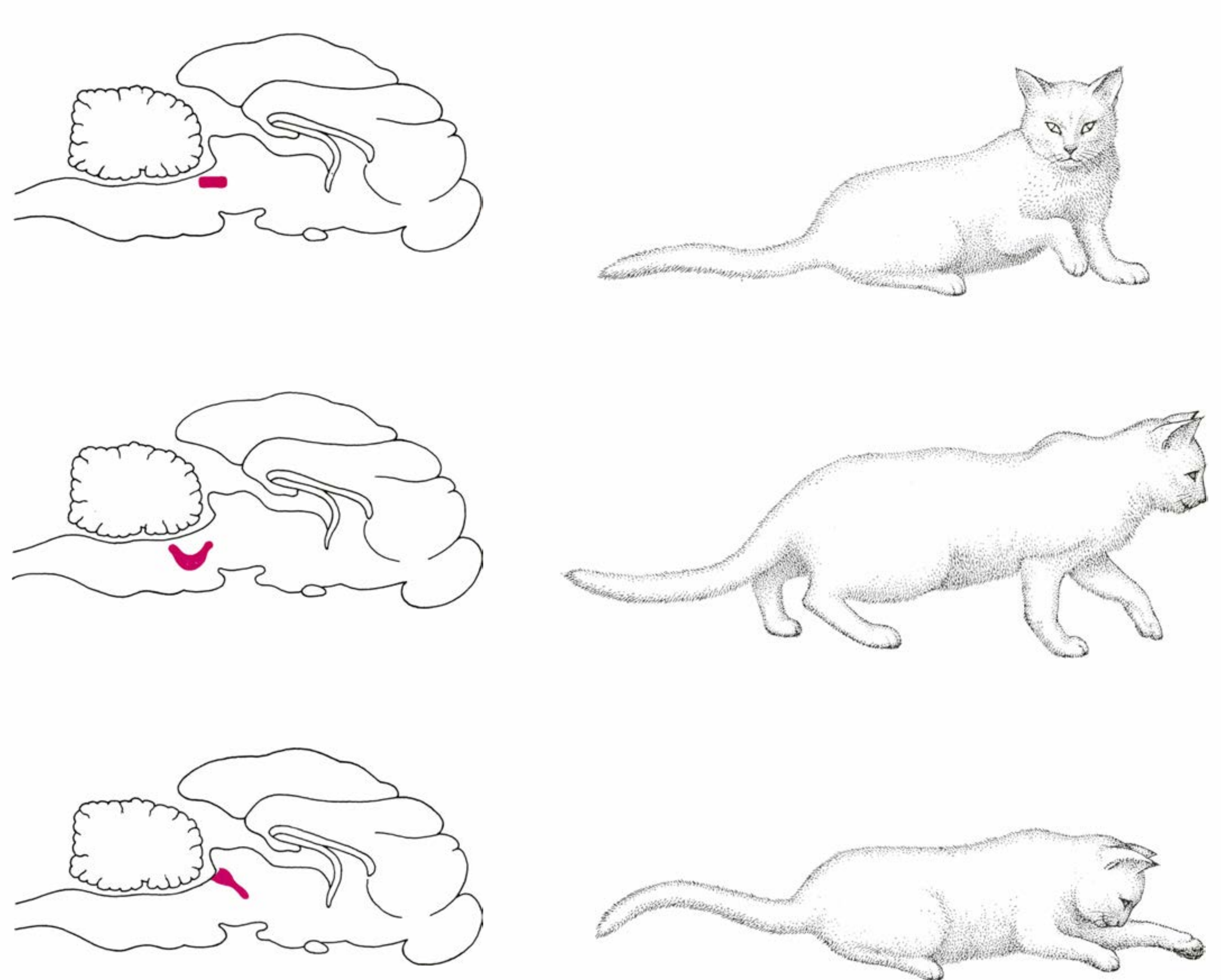
La investigación de ese eslabón podría tener, al menos, una consecuencia clínica importante. La enfermedad llamada narcolepsia se caracteriza por pasos súbitos e impredecibles, de la vigilia, directamente al sueño REM, o de la vigilia a la parálisis sin pérdida de la conciencia. Curiosamente, los estímulos fuertes, como los que acompañan a la ira, risa, sorpresa o la relación sexual, son las causas más comunes de las crisis narcolépticas. Puede que en quienes padecen narcolepsia el umbral que separa la vigilia de la inhibición motora sea demasiado bajo. Un estímulo co-

mún podría así conducir a la atonía o al sueño REM en situaciones inadecuadas.

Además de sus posibles aplicaciones clínicas, la investigación del sueño REM sin atonía probablemente nos informe de la naturaleza misma de la fase REM. Ha contribuido ya a esclarecer un problema fundamental. Como hemos visto, en la mayoría de las regiones del cerebro las neuronas siguen el mismo patrón en la transición del estado de vigilia al estado de sueño. Hay un descenso del ritmo de actividad entre la vigilia y el sueño lento y luego un aumento al iniciarse un episodio del sueño REM. Entre las neuronas que muestran una conducta diferente se cuentan las de la protuberancia conocida por núcleo dorsal del rafe. Esas células utilizan el transmisor serotonina para modificar la actividad de las neuronas a las que están conectadas. Su ritmo de disparo pasa de dos impulsos por segundo en vigilia a casi cero en el sueño REM, después de atravesar un nivel intermedio en el sueño de ondas lentas.

Se había sugerido que la inactividad de las neuronas del núcleo dorsal del rafe constituía una característica fundamental del sueño REM. De hecho, se había especulado con que el descenso de su actividad provocaba el comienzo del sueño REM. Para contrastar estas hipótesis me puse en comunicación con Barry L. Jacobs y Michael Trulson, de la Universidad de Princeton, quienes habían estudiado a fondo la actividad de las neuronas del rafe dorsal en el estado de vigilia y en el sueño normal. Al registrar la actividad de esas células en el sueño REM sin atonía, los resultados nos sorprendieron. En los gatos con lesiones en la protuberancia, las neuronas del núcleo del rafe, en lugar de cesar la actividad, la aumentaban después del sueño de ondas lentas. El incremento situaba la frecuencia de disparo en un impulso por segundo, esto es, no alcanzaba el nivel de vigilia; sin embargo, ese valor era unas seis veces superior al del sueño REM normal.

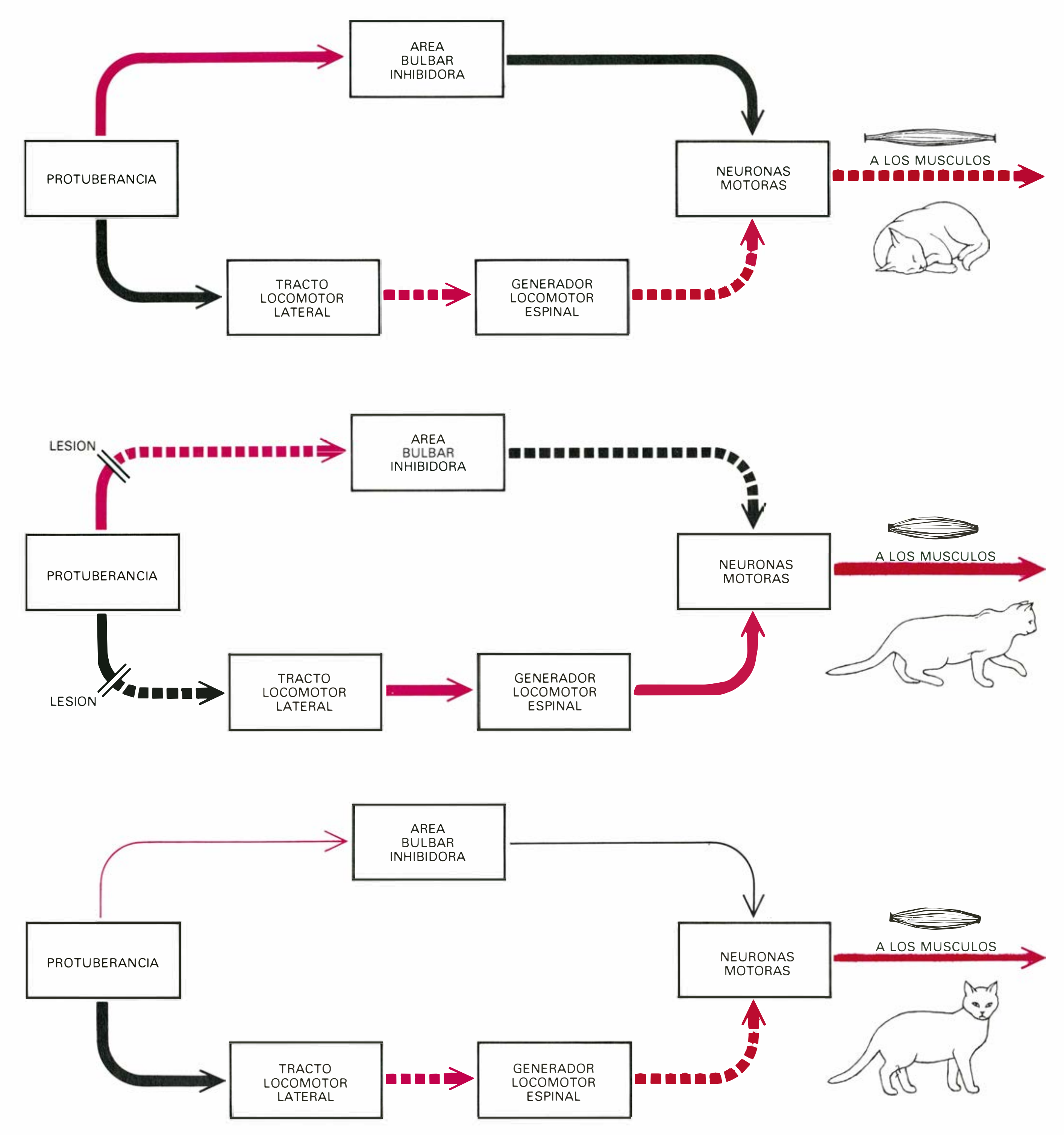
Existen, al menos, dos explicaciones plausibles de ese inesperado descubrimiento. De acuerdo con la primera, la actividad muscular del sueño REM sin atonía provoca el retorno de información a la protuberancia para que se exciten las neuronas del núcleo dorsal del rafe. La segunda explicación razona que las lesiones de la protuberancia afectan a un mecanismo motor más central, que normalmente inhibe las neuronas del rafe en el sueño REM. La primera hipótesis depende del envío



7. LA POSICION DE LAS LESIONES en el cerebro del gato afecta intensamente el tipo de movimientos que muestra el animal en el sueño REM sin atonía. Un gato con una pequeña lesión en la parte alta de la protuberancia sólo logra mantenerse sobre las patas delanteras (*arriba*). Tal lesión interrumpe únicamente parte del efecto que ejerce la zona inhibidora de la línea media sobre el tracto locomotor mostrado en la ilustración precedente. Un gato con una lesión mayor y más profunda se mantiene sobre las cuatro patas y camina (*centro*), pues la lesión bloquea completamente el efecto de la zona inhibidora. Un gato con una lesión que alcanza al mesencéfalo muestra una conducta agresiva (*abajo*). El daño afecta las vías que regulan la conducta agresiva.

de impulsos de la periferia del sistema nervioso a sus centros, mientras que la segunda se apoya tan sólo en fenómenos cerebrales.

En la comprobación de las hipótesis, Jacobs y sus ayudantes emplearon dos drogas. Por un lado, succinilcolina, una droga químicamente muy afín al curare, que actúa en la unión entre una neurona motora y la célula muscular que inerva. Al interrumpir la conexión, la droga induce una parálisis temporal. Cuando se inyectó succinilcolina a gatos normales despiertos, las neuronas del rafe no modificaban su actividad. Las neuronas del rafe podían mantener, pues, la actividad en la vigilia con parálisis y en el sueño REM sin atonía y, por tanto, sin parálisis. Puesto que las neuronas del rafe se mostraban activas en animales paralizados en vigilia, concluimos que las neuronas del núcleo del rafe no recibían el estímulo, en el sueño REM sin atonía, de información



8. LOS SISTEMAS QUE INHIBEN EL MOVIMIENTO en el sueño REM actúan también en situaciones de emergencia durante la vigilia. Se muestran, de forma esquemática, en las tres partes de la ilustración. Las conexiones excitadoras aparecen en color y, las conexiones inhibidoras, en negro. En el sueño REM normal, la protuberancia activa intensamente el centro inhibidor del bulbo (*arriba*). La zona media inhibidora de la protuberancia inhibe el tracto locomotor lateral. Resulta de ello una parálisis completa. En el sueño REM sin parálisis, las lesiones interrumpen las conexiones que unen la protuberancia con el tracto locomotor y con el centro bulbar (*centro*). El gato se mueve, pero la liberación del tono muscular no suele ser completa. En vigilia, cuando se percibe un estímulo inesperado de amenaza, se produce a veces una inhibición transitoria del movimiento (*abajo*). La inhibición temporal probablemente sea consecuencia de la reducción de la conducción motora, puesto que la pérdida de tono muscular es muy leve. Dibujo de Tom Prentiss.

transmitida desde la periferia, resultante de una actividad muscular vigorosa.

La segunda droga, el carbacol, causa parálisis al afectar directamente al mecanismo responsable de la atonía descrito antes. La inyección de carbacol silenció las neuronas del rafe casi por completo. Más que una parte fundamental del sueño REM, o un elemento del mecanismo que induce un episodio de sueño REM, la inactividad de las neuronas del rafe dorsal semeja, pues, un fenómeno secundario. Parece guardar relación con la inhibición motora central del sueño REM, que puede invertirse por lesión de la protuberancia. Mi ayudante Peter Reiner ha abordado el estudio del otro grupo de células nerviosas cuya actividad se suspende en el sueño REM; se encuentran cerca del lugar donde practicamos nosotros las lesiones y emplean noradrenalina como transmisor. Los primeros resultados obtenidos por Reiner sugieren que la inactividad de esas células guarda también relación con la inhibición motora.

Mediante técnicas experimentales que nos permiten separar la actividad cerebral de la parálisis muscular asociada al sueño profundo estamos recogiendo información sobre lo que es fundamental en el sueño REM y lo que es un efecto colateral de uno de esos episodios. La aceptación de las hipótesis en boga sobre la parálisis en el sueño REM requerirá, sin duda, un gran esfuerzo investigador, en particular la idea, adelantada antes, de la existencia de vías separadas para la inhibición del tono muscular y de la conducción motora, así como la hipótesis de que el segundo trayecto sigue activo en la vigilia.

Sean o no acertadas esas ideas, el sueño REM sin atonía permanecerá como valiosa técnica experimental. Por añadidura, separar lo esencial de lo secundario en esa fase podría dotarnos de un medio de estudio de conductas complejas, tal la agresión, que suelen producirse en respuesta a estímulos externos, aunque en el sueño REM podrían generarse con independencia del entorno del animal. Además, y dadas las similitudes entre la vigilia y el sueño REM, la comparación de ambos estados siempre será de interés. En los movimientos que ejecuta un animal despierto tenemos una conducta correlacionada con su estado interior. Nos faltaba una correlación similar para el sueño REM; el sueño REM sin atonía nos la proporciona, siendo, en cierto sentido, una ventana abierta al cerebro dormido.

Primeros agricultores de la llanura del norte de Europa

Excavaciones efectuadas en Polonia han puesto al descubierto los restos de pobladores que cultivaban, pastoreaban y cazaban en aquella zona hace 7000 años. Se trataba de un tiempo de transición hacia la agricultura

Peter Bogucki y Ryszard Grygiel

Uno de los más importantes acontecimientos de la prehistoria humana –el surgimiento de la agricultura y la ganadería– ocurrió en momentos y lugares diferentes del Mundo Antiguo. El conocimiento de ese nuevo modo de vida en el sur de Europa fue, quizás, el resultado de un paulatino proceso de difusión desde aquellas zonas del Próximo Oriente donde la revolución neolítica se había producido de un modo más precoz. Sin embargo, en la Europa situada al norte de los Cárpatos y los Alpes, el avance neolítico fue un movimiento de gentes, más que de ideas; el registro arqueológico nos documenta claramente alguno de esos movimientos migratorios.

Tras varias campañas de trabajos de campo, el conocimiento que poseemos sobre los pobladores que abandonaron el valle medio del Danubio, llevando consigo plantas y animales domesticados, para internarse en la llanura forestal norteeuropea, dista mucho de ser satisfactorio. La falta de datos se debe, por una parte, al hecho de que la acidez del suelo del norte destruye fácilmente los restos humanos y animales, y, de otra, a la preferencia de los arqueólogos por realizar sus trabajos en los extensos yacimientos del Neolítico Final de la llanura polaca, en detrimento de los del Neolítico Inicial. A pesar de todo, en la década de 1930, los yacimientos del Neolítico Inicial de la llanura polaca, situados a unos 150 kilómetros al noroeste de Varsovia, comenzaron a arrojar valiosa información, hasta que la segunda guerra mundial detuvo los trabajos. Algunos yacimientos continúan hoy aportándonos datos. Pero antes de pasar a describirlos, convendrá familiarizar al lector con el medio físico de ese amplio cinturón transmontano y con las oportunidades que ofrecía a los primeros emigrados del sur.

En sentido lato, Europa central puede dividirse en tres grandes zonas ecológicas. La más meridional de ellas, la zona montañosa de los Cárpatos y los Alpes, no fue, por lo general, lugar de asentamiento para las comunidades agrícolas hasta bien entrado el Neolítico. Por tanto, ese área no será objeto de nuestro estudio aquí. Al norte, más allá de las montañas, nos encontramos con una altiplanicie que se extiende desde Checoslovaquia y el sur de Polonia hasta Alemania central y Francia. La región se encuentra ampliamente cubierta por loess, la fértil y seca capa de tierra depositada por los vientos que soplaban a través de Europa en la última etapa glaciar. Los loess rellenan las cuencas de los ríos de la región y cubren las colinas adyacentes, alcanzando en algunos lugares una profundidad de 100 metros.

La tercera área, la más septentrional, abarca las tierras bajas de la llanura norteeuropea, extendiéndose desde el este de la Unión Soviética, a través de Polonia, norte de Alemania y Dinamarca, hasta los Países Bajos. Las capas de hielo de la última glaciación cubrieron esa zona hasta hace unos 15.000 años; su avance excavó pequeños lagos y su retroceso dejó un terreno pantanoso, con hondonadas llenas de agua y un suelo cubierto de arena, arcillas y barros. Algunos de estos nuevos suelos llegaron a ser muy fértiles; otros permanecieron estériles. Todos ellos, sin embargo, sustentaban una capa forestal diversa y productiva, y los lagos y corrientes de agua eran ricos en vida acuática.

Los agricultores procedentes del sur que penetraron en las tierras altas se desplazaron en primer término hacia el noroeste, a través de los ríos más importantes de Europa: Danubio, Elba, Main y Rhin. Se asentaron en la región de las tierras altas más ricas, donde los loess cubrían las cuencas de los ríos y donde, no obstante, las plantas silvestres y animales salvajes eran relativamente escasos. Las comunidades neolíticas recientemente establecidas dependían rigurosamente de los animales que llevaban consigo (reses, ovejas, cabras y cerdos) y de las plantas (cebada y trigo). Con el tiempo, algunos inmigrantes avanzaron más al norte, a través del Vístula y el Oder, hacia la llanura norteeuropea.

Allí se encontraron con tierras y suelos notoriamente diferentes de aquellos de las tierras altas. El nuevo ecosistema ofrecía más de 450 especies de plantas comestibles; pesca, tortugas y moluscos en los ríos y lagos; ciervos comunes (*Cervus elaphus*), corzos (*Capreolus capreolus*), reses salvajes, jabalíes, e incluso caballos salvajes, en los bosques. Esa zona de excepcional productividad sólo presentaba un inconveniente. Sus rendimientos no eran constantes a lo largo del año, pues durante los meses invernales mermaban notablemente. La adaptación de estos primeros neolíticos a tan desigual productividad constituye una de las cuestiones más importantes que quizá puedan resolver los datos aportados por los yacimientos polacos.

De un total de más de 20 yacimientos neolíticos conocidos ahora en la vecindad de Brześć Kujawski, una ciudad del distrito polaco de Wloclawek, dos (identificados con la numeración 3 y 4) se descubrieron en 1933; cuando unos agricultores extraían grava a la orilla de un lago seco, aparecieron objetos y huesos humanos. Konrad Jażdżewski, del Museo Arqueológico Nacional de Varsovia, tuvo noticia del descubrimiento y realizó siete campañas (1933-39) de excavación en los yaci-

1. TRAZADO DE LAS PAREDES de una casa alargada, puesto al descubierto tras eliminar la capa del suelo de labranza. Guiaba la excavación de un asentamiento del primer neolítico en Brześć Kujawski a medida que los investigadores reseguían la arcilla descolorida donde había estado clavada la madera de la estructura. Manchas similares del subsuelo indicaban la presencia de hoyos de desecho, rellenados con objetos y huesos de animales. La excavación irregular que aparece en primer término condujo a una capa de arcilla que los ocupantes neolíticos usaban para revocar las paredes de sus casas y fabricar cerámica. Este es el más tardío de los asentamientos principales, ocupado por el grupo de la cultura Lengyel. La datación por carbono 14 calibrado fija su llegada entre el 4500 y el 4400 a. de C. y, el abandono del asentamiento, hacia el 3900 a. de C. Fotografía tomada por Grygiel.



2. TRES ENTERRAMIENTOS aparecieron asociados a la casa alargada de la fotografía superior. El esqueleto de la izquierda, de un hombre, yace tendido sobre su lado derecho; los otros dos, de mujeres, aparecen recostados sobre su lado izquierdo. Los restos están orientados de forma que los cráneos apuntan al sur-sureste. Se han excavado más de 70 tumbas de los Lengyel, pero no ha aparecido ninguna de los primeros ocupantes de los yacimientos 3 y 4, de la cultura de la cerámica lineal. Muchos enterramientos de Lengyel van acompañados de ricos objetos funerarios, ornamentos de cobre incluidos.



3. CUCHARA DE ASTA DE CIERVO procedente de uno de los hoyos de desecho del yacimiento número cuatro. Es el primer objeto de ese tipo hallado entre los restos neolíticos de Europa central. Los Lengyel también obtenían picos de las astas de esos animales; las recogían aprovechándose de la muda.

mientos, llegando a investigar un área total de más de 10.000 metros cuadrados. Su trabajo dio a conocer muchas estructuras de casas neolíticas, más de 50 enterramientos humanos, alguno de ellos ricamente provisto de objetos de cobre, hueso y ornamentos, como collares de cuentas, así como picos de asta de ciervo. Tanto los huesos humanos como los de los animales se habían conservado excelentemente en aquel suelo no ácido. Los objetos de cobre fueron los primeros en aparecer en esa parte de Europa central; el metal procedía seguramente de las minas de los Cárpatos meridionales. La guerra detuvo la investigación de Jażdżewski y los objetos recogidos en la campaña de 1939 quedaron destruidos al sufrir un bombardeo las cocheras de tranvías de Varsovia. Su trabajo, no obstante, había probado la importancia de tales yacimientos para la comprensión de la vida de los primeros neolíticos de la llanura norteeuropea.

Tanto el trabajo de Jażdżewski, realizado antes de disponerse del método de datación por carbono 14, como la posterior investigación de la zona, dejaron sin respuesta una serie de cuestiones cronológicas y económicas. De ahí que uno de nosotros (Grygiel), del Museo Arqueológico y Etnográfico de Lódź, en compañía del otro (Bogucki), por entonces un estudiante de licenciatura en la Universidad de Harvard, hiciéramos un reconocimiento, en el verano de 1976, de los yacimientos números 3 y 4. Observamos que en ambos yacimientos quedaban aún importantes zonas por excavar. Con ayuda financiera del distrito de Wloclawek y de la ciudad de Brześć Kujawski, hemos llevado a cabo siete campañas de excavación, realizadas, en su mayor parte, en el yacimiento número 4.

Como en la mayor parte de los yacimientos neolíticos de la Europa templada, los restos arqueológicos están aquí a una profundidad relativamente pequeña. El primer estrato, formado por la llamada tierra negra, descansa sobre un substrato de arena y arcilla. Las labores de arado remueven cada año los sedimentos hasta una profundidad de más de 30 centímetros. El método de excavación, por tanto, consistió en eliminar la zona arada, exponiendo así cualquier característica impresa en la arcilla y grava subyacentes. A diferencia de la acumulación seriada de depósitos, típicos del yacimiento muy estratificado, lo que aparece a la vista es un palimpsesto de hoyos, fosas, zanjas y tumbas, frecuentemente superpuestos. Tales superposiciones permiten determinar una cronología relativa de los sucesivos asentamientos. La datación absoluta viene dada por los estudios radiocarbónicos, de los que se han efectuado un total de 15, en laboratorios de Polonia, Estados Unidos y Países Bajos.

En los yacimientos se evidencian dos períodos principales de ocupación neolítica, separados por un intervalo de dos o tres siglos, en que el asentamiento fue de menor entidad. El período inicial comenzó hacia el 5300 a. de C. (la fecha verdadera según la datación cronológica por carbono 14 calibrado) y se prolongó aproximadamente hasta el 4800 a. de C. Los ocupantes pertenecían a un grupo que los arqueólogos europeos denominan cultura de la cerámica lineal, debido a la típica decoración de líneas incisas de su cerámica.

Restos de esa cultura se encuentran por toda Europa, desde Ucrania hasta Francia, tanto en las tierras altas de loess como en parte de la llanura norteeuropea. Muchos de los yacimientos de las tierras altas nos muestran vestigios de casas alargadas, algunas veces una simple granja, pero más comúnmente verdaderas viviendas agrupadas en poblados. No se han hallado casas alargadas de la cultura de la cerámica lineal en la llanura norteeuropea, y los dos yacimientos que hemos excavado tampoco han sido la excepción. Entre los hoyos que indicaban la presencia de cerámica lineal, amplios pero poco profundos, había dispersos algunos agujeros de postes, que probablemente señalaran la localización de cobertizos o similares refugios de temporada. Se pueden apreciar dos fases dentro de la

4. CUENTAS DE CONCHA ensartadas en collares, descubiertas entre los objetos de una tumba asociada a un esqueleto femenino de Lengyel. En un enterramiento se hallaron más de 5000 cuentas de ese tipo.

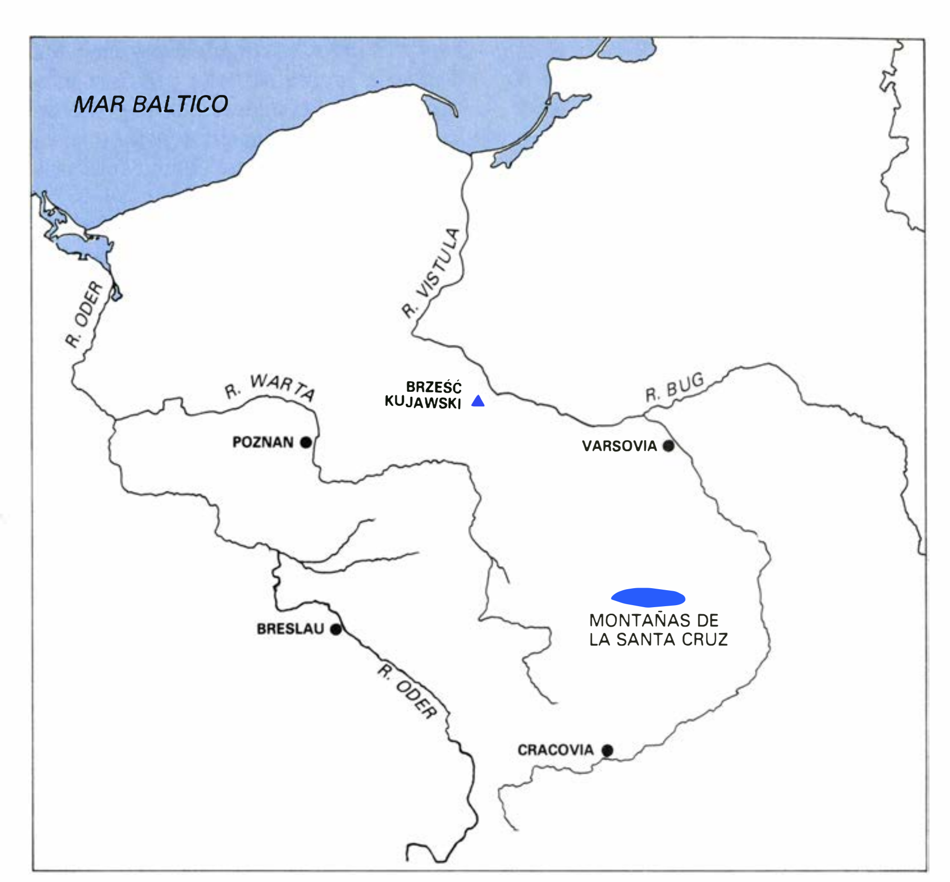
ocupación de la cerámica lineal. La primera de ellas, en el yacimiento número 3, se asentó en el extremo de una península, a la orilla de un lago; la segunda se situó a unos 200 metros, ya en el yacimiento número 4. En ninguno de los dos se encontraron enterramientos con cerámica lineal.

Siglos después, inmigrantes asociados por los arqueólogos europeos a la cultura de Lengyel alcanzaron esta zona y ocuparon ambos asentamientos. Esta temprana manifestación neolítica, que recibe el nombre de su original asentamiento en Hungría, representa, en esencia, una continuación de las tradiciones de la cerámica lineal del Este al centro de Europa; se caracteriza, empero, por nuevas formas cerámicas, así como por un declive de gran parte de su decoración. Muchos de los yacimientos Lengyel tienen también casas alargadas de planta trapezoidal, siendo marcadamente más estrechas en un extremo que en otro. Casas alargadas de esta peculiar planta se han descubierto en las tierras altas de loess y en la llanura del norte de Europa; son los primeros abrigos permanentes que aparecen en ambos yacimientos de Brześć Kujawski.

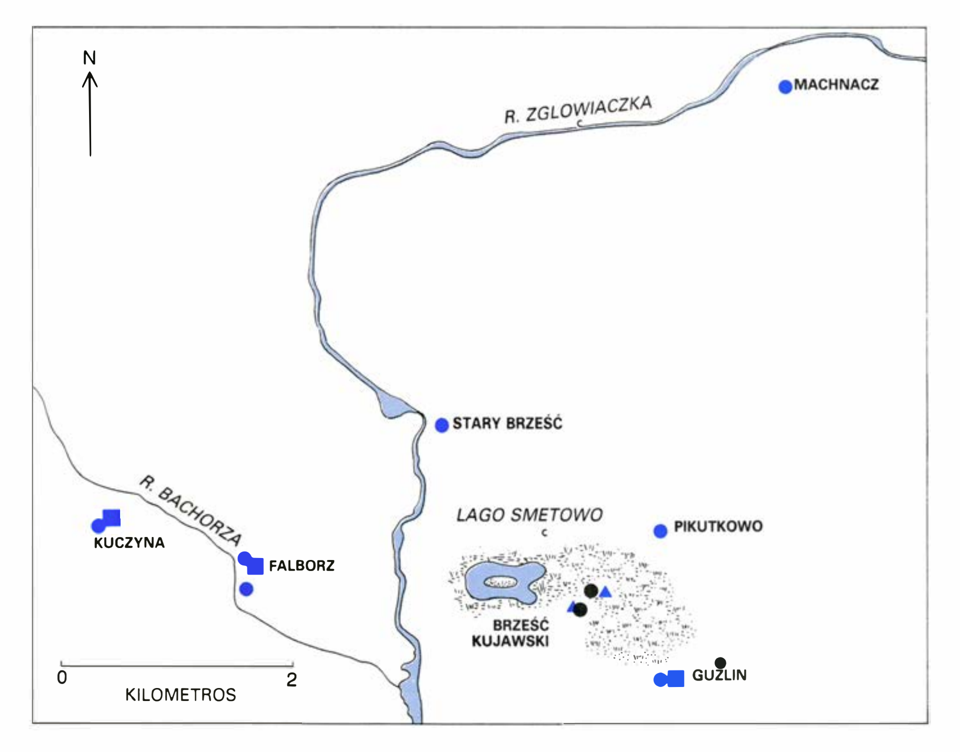
En la cultura de Lengyel se distinguen cuatro fases: la más temprana, visible en el yacimiento número 4, aparece entre el 4500 y el 4400 a. de C., y se prolonga hasta cerca del 4300 a. de C. Las tres fases posteriores, que abarcan desde el 4300 hasta cerca del 3900 a. de C., son patentes en ambos yacimientos. En la década de 1930, Jażdżwski denominó, a estas tardías variantes locales de la cultura Lengyel, grupo de Brześć Kujawski, y así se le sigue llamando todavía.

El rasgo más sobresaliente de ambos yacimientos son los vestigios de las casas alargadas de Lengyel. Aparecen en forma de manchas oscuras en la arcilla, la huella de las zanjas excavadas para la sustentación de las paredes de los habitáculos. Desde el inicio de los trabajos, en la década de 1930, hasta hoy se han descubierto más de 50 plantas de casas, cuya longitud supera, en algún caso, los 30 metros. Todas ellas tienen orientados sus ejes de noroeste, el lado más estrecho, a sudeste, el más ancho. Por término medio, las casas alargadas de Lengyel tenían 20 metros de longitud; el lado noroeste medía cerca de tres metros de ancho y, el sudeste, unos cinco metros.

No se sabe a ciencia cierta por qué la gente de Lengyel levantaba refugios trapezoidales. ¿Debíase, acaso, a que el lado estrecho mirase cara a los vien-



5. YACIMIENTOS DE BRZEŚĆ KUJAWSKI, señalados por el triángulo de color. Se encuentran a unos 150 kilómetros al oeste-noroeste de Varsovia, y casi a 200 kilómetros de las montañas de la Santa Cruz, donde abunda el sílex achocolatado, así llamado por su peculiar coloración. Este singular sílex era muy apreciado por el grupo de la cerámica lineal debido a sus excelentes propiedades para la talla. Los posteriores pobladores se contentaban con el sílex, de peor calidad, depositado por las glaciaciones.



6. LAS SUCESIVAS OCUPACIONES de Brześć Kujawski se erigieron en una península que se proyectaba al lago Smetowo, entonces de gran tamaño. Las ocupaciones del grupo de la cerámica lineal se realizaron cerca del extremo peninsular y de su base (*puntos negros*); las ocupaciones Lengyel en esos mismos asentamientos aparecen con triángulos de color. Otros asentamientos del grupo de la cerámica lineal (*puntos de color*) se han hallado al sur, al este y al norte. También se han descubierto tres asentamientos temporales (*cuadrados de color*). Uno de ellos, a juzgar por la cantidad de picos que se encontraron, fue un aserradero o campamento donde se trabajaba la madera. Los dibujos son de Alan D. Iselin.

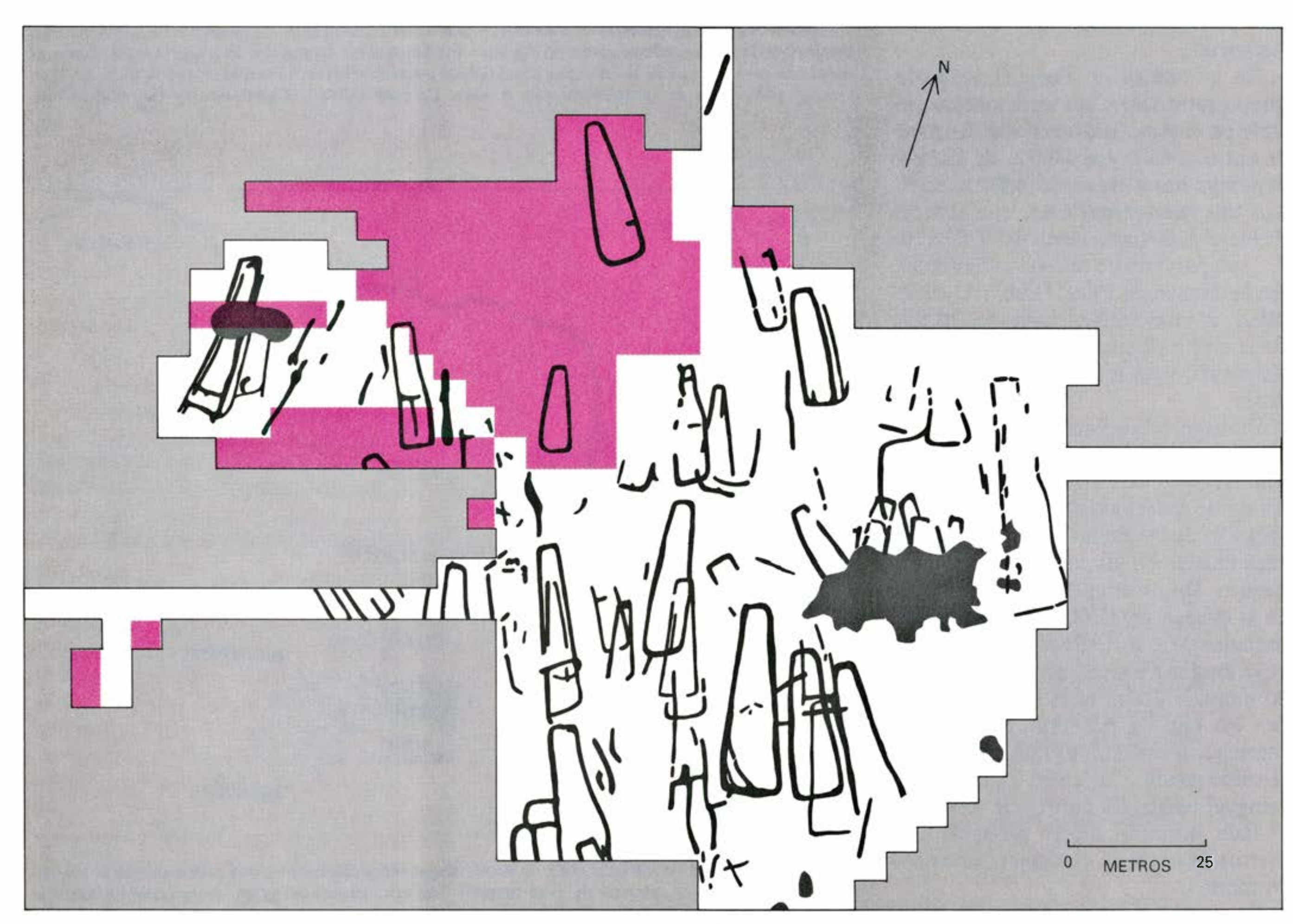
tos predominantes? En 1976 se erigió en Francia, para la televisión, una réplica de una casa alargada y simétrica. La cara expuesta a los vientos dominantes ha sufrido ya serios daños, mientras que las otras tres permanecen intactas. En los yacimientos se aprecian varios ejemplos de reconstrucciones de casas alargadas de los Lengyel. La reutilización de algunos asentamientos sugiere que la ocupación no cesó en las cuatro fases; probablemente hubiera unas diez casas ocupadas en todo momento.

Los rasgos más comunes y significativos que subyacen a los dos asentamientos son los hoyos, de variado tamaño y formas. Como hemos señalado, los hoyos asociados a las gentes de la cerámica lineal son superficiales y anchos. Posiblemente en el tiempo de ocupación no fueran más que hoyos naturales de la tierra, como el que deja un árbol arrancado. Contienen una densa acumulación de escombros, formados por huesos de animales y piedras de desecho, sílex y cerámicas. Por el contrario, muchos de los hoyos de Lengyel son profundos e irregulares, excavados donde la arcilla se halla próxima a la superficie. Ello sugiere que la gente de Lengyel los cavaba en principio para obtener arcilla con que enlucir sus casas y fabricar cerámica.

Otros de los hoyos de Lengyel suelen ser profundos y circulares, y se han encontrado en la orilla de lo que anteriormente era un lago. El cieno depositado en los fondos de los hoyos sugiere que debieron utilizarse para conservar, en agua, moluscos, tortugas y peces vivos. En todo caso, la mayoría de los pozos de extracción de arcilla o de conservación de animales acababan empleándose para la deposición de basuras y desechos. A diferencia de los hoyos de la cerámica lineal, los de Lengyel muestran una considerable variedad de desechos. Los huesos de animales y los objetos de desecho abundan más en los hoyos próximos a las casas.

En algunas zonas donde se removió el subsuelo en tiempos neolíticos, la decoloración es menor que en los agujeros de desechos; al parecer, el área que había sido excavada no se rellenó con elemento orgánico alguno. Tras su investigación, tales áreas parecían corresponder, generalmente, a enterramientos de los Lengyel. Se han excavado ya más de 70 enterramientos Lengyel, el más importante de ellos por Jażdżewski, en la década de 1930. Los esqueletos solían encontrarse en este tipo de fosas individuales. Por lo general aparecen contraídos, con las rodillas flexionadas hacia arriba; los hombres se enterraban acostados sobre su lado derecho y las mujeres sobre el izquierdo. En ambos sexos la orientación horizontal era la misma: los cráneos se situaban en el extremo sudeste de la fosa. Las mujeres se enterraban a menudo adornadas con collares de cuentas y otros ornamentos de concha, hueso y cobre; por otra parte, los picos de asta de ciervo son frecuentes en los enterramientos masculinos. No se han hallado enterramientos del período de cerámica lineal.

El conjunto de objetos abandonados tanto por las gentes de Lengyel como por las de la cerámica lineal incluyen los típicos picos y azuelas talladas y pulimentadas del Neolítico. En efecto, el pico de este tipo (conocido



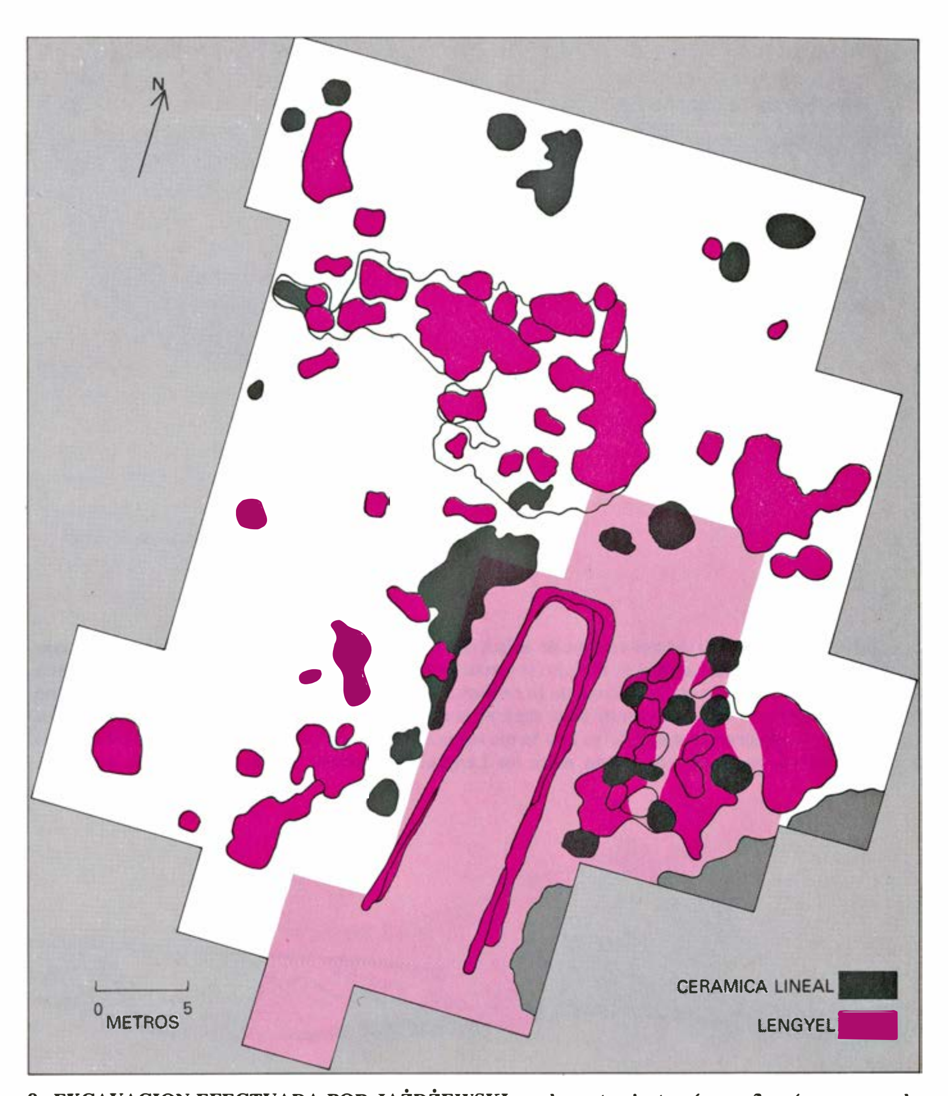
7. MAS DE 50 RASTROS de paredes de casas alargadas, muchos de ellos superpuestos, excavó en el yacimiento número 4 Konrad Jażdżeswki, del Museo Arqueológico Nacional de Varsovia, entre 1933 y 1939. Los trazados de casas que aparecen en el área blanca corresponden al trabajo realizado en la década de 1930. Las zonas de color gris oscuro indican las demoliciones modernas anteriores a las excavaciones de Jażdżewski. Las áreas de color fueron investigadas por los autores del artículo entre 1976 y 1982. Por mor de claridad no se han representado los enterramientos y los hoyos de desecho.

como pico celta) es característico de la cultura de la cerámica lineal. Estos primeros agricultores fabricaban todavía pequeños instrumentos de sílex. No obstante, el trabajo del sílex difería bastante en ambas culturas. En la de la cerámica lineal se prefería una especialidad polaca: el sílex achocolatado, así llamado por su coloración marrón y por encontrarse únicamente en las montañas de la Santa Cruz, a unos 200 kilómetros al sudeste de los yacimientos. El sílex achocolatado polaco era muy valorado por los fabricantes de utensilios de piedra incluso en tiempos paleolíticos. Posee unas excelentes cualidades para ser sometido a la talla, y las gentes de la cerámica lineal lo utilizaron con profusión para obtener finas hojas de sílex.

En la cultura Lengyel no se aprecia una preferencia tan marcada. Se fabricaba la mayor parte de los utensilios de sílex a partir del abundante sílex "errático", depositado en todo el norte de Europa por el retroceso de los hielos. En gran parte, el sílex errático había estado sometido a presión y otras tensiones durante su desplazamiento, lo que mermó sus propiedades para la talla; así se aprecia en sus láminas. Sólo cuando necesitaban fabricar un útil de precisión se tomaban la molestia de procurarse y trabajar el sílex achocolatado.

Los Lengyel fueron un paso más lejos que los de la cerámica lineal en la talla y pulido de picos. Practicaban una perforación para su enmangue; en los hoyos de desechos se encuentran los fragmentos extraídos en el procedimiento. Las gentes de Lengyel también recogían las astas de los ciervos, tanto para hacer picos como para fabricar otros utensilios. El más notable de ellos es una cuchara de asta de ciervo, única en Europa central, que descubrimos el año pasado.

Los huesos y dientes de animales también servían para engalanar los collares y pendientes. Las costillas de los grandes mamíferos (probablemente reses) se curvaban para hacer amuletos, al parecer cuando aún estaban blandas, y luego se decoraban con motivos incisos. De la concha nacarada de un molusco de agua dulce (*Unio*) se fabricaban cuentas de adorno. Collares de este tipo ceñían la cintura de una hembra en un enterramiento en el que se encontraron más de 5000 conchas de *Unio*, cada una de un centímetro de diámetro. En las tumbas de hembras Lengyel se descubrieron también cuentas de cobre y amuletos; cintas y varillas de cobre dispersas por la fosa indican que esos ornamentos se obtenían en el mismo lugar a partir de mena de cobre alóctona.

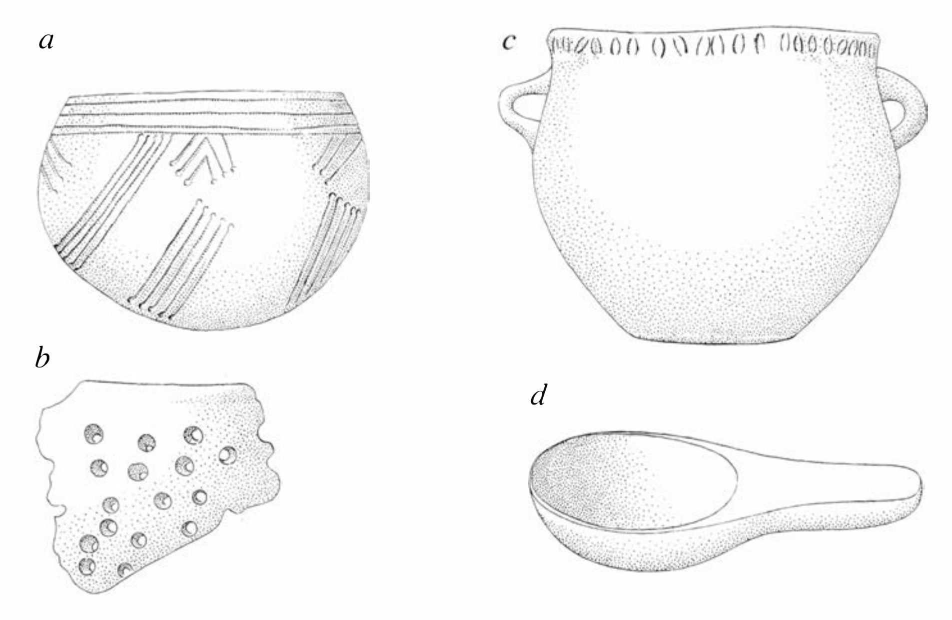


8. EXCAVACION EFECTUADA POR JAŻDŻEWSKI en el asentamiento número 3, más cercano a la orilla del lago. Corresponde a la parte tramada en color en este plano. Se aprecian los trazados de dos casas alargadas casi perfectamente superpuestas y un número de hoyos de desecho, similarmente coincidentes, que contenían objetos y restos de animales pertenecientes tanto a los ocupantes de la cerámica lineal como a los del grupo de Lengyel. Es patente la planta trapezoidal de las casas alargadas de Lengyel; el lado más estrecho probablemente se encarara a los vientos predominantes. Aunque se conocen casas alargadas de la cultura de la cerámica lineal en todo el resto de Europa, no se ha descubierto ninguna en la llanura norteeuropea. Los agujeros de los postes asociados a los objetos de la cerámica lineal sugieren que estos primeros emigrantes del Neolítico construían exclusivamente abrigos de empleo temporal.

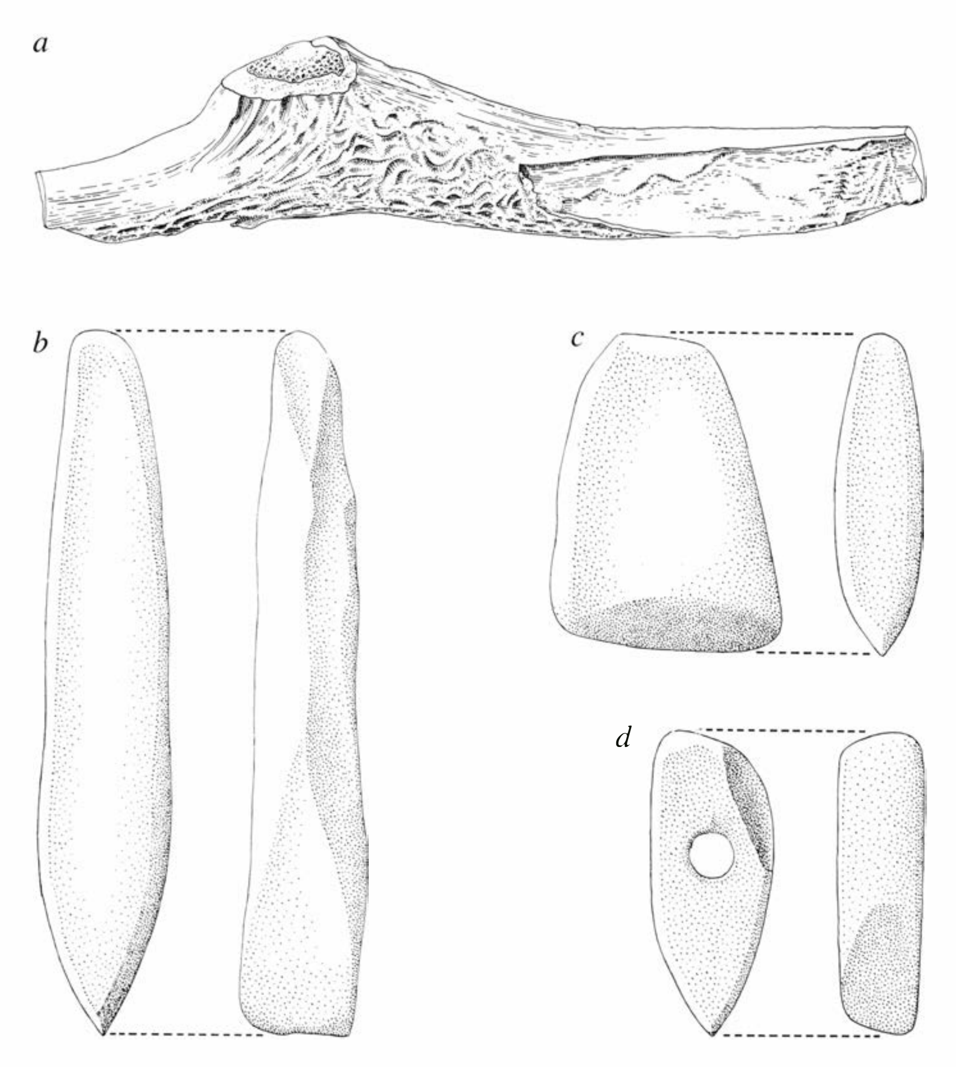
La cerámica que da nombre a la cultura lineal difiere de la cerámica Lengyel por la decoración y por sus formas. Las más comunes en Brześć Kujawski coinciden con las halladas en otros yacimientos de cerámica lineal en Europa central: cuencos profundos de tres cuartos de esfera. Otra de las formas cerámicas común en ambos yacimientos polacos y otros lugares de la llanura centroeuropea, pero relativamente escasa en las tierras altas, es el colador. No contamos con coladores íntegros, pero el número de fragmentos hallados sugiere que servían para separar la cuajada del suero en la fabricación de queso. Robert Cowie, de la Universidad de Sheffield, viene haciendo estudios de fragmentos de cerámica lineal para averiguar si su composición mineral puede servir para determinar el yacimiento originario de las arcillas empleadas en su manufactura.

La cerámica de los Lengyel posee formas que van desde los cuencos poco profundos hasta vasijas como las ánforas griegas, y cuenta con artículos especializados, como las cucharas de barro. Sus componentes minerales no se han sometido a análisis detallados, pero la cantidad de mica que contienen las vasijas sugiere que se fabricaron con la materia prima disponible en la localidad. Eso mismo se deduce de que la arcilla se extrajera de los hoyos del asentamiento.

En nuestros trabajos, iniciados en 1976, hemos recogido gran cantidad de huesos de animales en buen estado de conservación. Por ejemplo, hemos extraído más de 6000 espinas, la



9. CUENCO PROFUNDO de tres cuartos de esfera (*a*), ejemplo del estilo decorativo lineal que da nombre a la cultura de la cerámica lineal. Debajo (*b*) aparece un fragmento de colador de barro, objeto común en los yacimientos polacos de la cultura de la cerámica lineal, aunque no se haya encontrado ninguno intacto. Los coladores debían de servir para separar la cuajada del suero en la fabricación de queso. La vasija con asas y decoración simple (*c*) es una forma típica de la cultura Lengyel. Debajo (*d*) se esboza otro de los objetos de barro de uso común entre los Lengyel: la cuchara. No se ha respetado la escala.



10 PICOS fabricados por talla o pulimento de la piedra o a partir de astas de ciervo (*a*). El pico de piedra de la izquierda (*b*), conocido como pico celta, es un útil característico de la cultura de la cerámica lineal. El más corto y ancho (*c*) es una forma característica de la cultura Lengyel. Los Lengyel también practicaban perforaciones en los picos para enmangarlos (*d*). No se ha encontrado en los yacimientos 3 y 4 ningún pico perforado; sí han aparecido los fragmentos de material extraídos en el proceso de su fabricación.

mayor parte de sargo y perca. De nuevo se aprecia un fuerte contraste entre los animales de los primeros ocupantes y los característicos de los ocupantes posteriores. Como ocurre en otros yacimientos de la cultura de la llanura norteeuropea, los primeros animales domesticados de la cerámica lineal eran ganado vacuno. Toda vez que muchos de los huesos hallados corresponden al extremo inferior de la gama de tamaños, hemos de pensar que los rebaños estarían constituidos fundamentalmente por reses. También serían objeto de pastoreo algunas ovejas y cabras; sin embargo, aunque estos agricultores mantenían muchos cerdos en los asentamientos de las tierras altas, sólo hemos encontrado indicios de tres ejemplares en los 500 años de ocupación de Brześć Kujawski. Destacan también por su pequeño número los restos de aves, peces, tortugas, moluscos y huesos de animales de caza, como el ciervo, corzo y el jabalí. Efectivamente, del elevado número de espinas de peces encontradas en ambos yacimientos, no llegan a 100 las que corresponden a la ocupación de la cerámica lineal.

No ocurre igual con el registro de animales asociados a la ocupación Lengyel. Aunque en general los restos de vacuno son más numerosos que los de la cerámica lineal, estos animales no figuran tan decisivamente en el conjunto de la cabaña. Ovejas y cabras son casi tan numerosas y los cerdos igualan en número al ganado vacuno; simultáneamente se observa un incremento de los animales producto de la caza: corzo, ciervo, aves acuáticas, tortugas, peces y moluscos. Pero no toda la caza se hacía con fines alimentarios. Por ejemplo, los castores se cazaban por sus pieles y lo mismo podía ocurrir con las nutrias, comadrejas y las liebres.

¿Qué queda del cuidado de los campos en estos asentamientos neolíticos tempranos? Hemos prestado especial atención a la recuperación de plantas por medio de técnicas como la criba pasada por agua y la flotación, pero hemos hallado nada más tres granos carbonizados en los pozos de desecho de la cerámica lineal, lo cual induce a pensar que esa gente se dedicaba muy poco al cultivo de cereales. Por el contrario, los hoyos de desechos de los Lengyel nos aportan gran cantidad de granos carbonizados. Uno de los hoyos excavados el último año contenía más de 300 muestras. Los primeros análisis, realizados por Caroline Quillian Stubbs, que trabajaba entonces en la Universidad de Tübingen, indicaron que la mayoría de los granos pertene-

cían a la forma primitiva de trigo conocida por los botánicos como emmer (*Triticum dicoccum*).

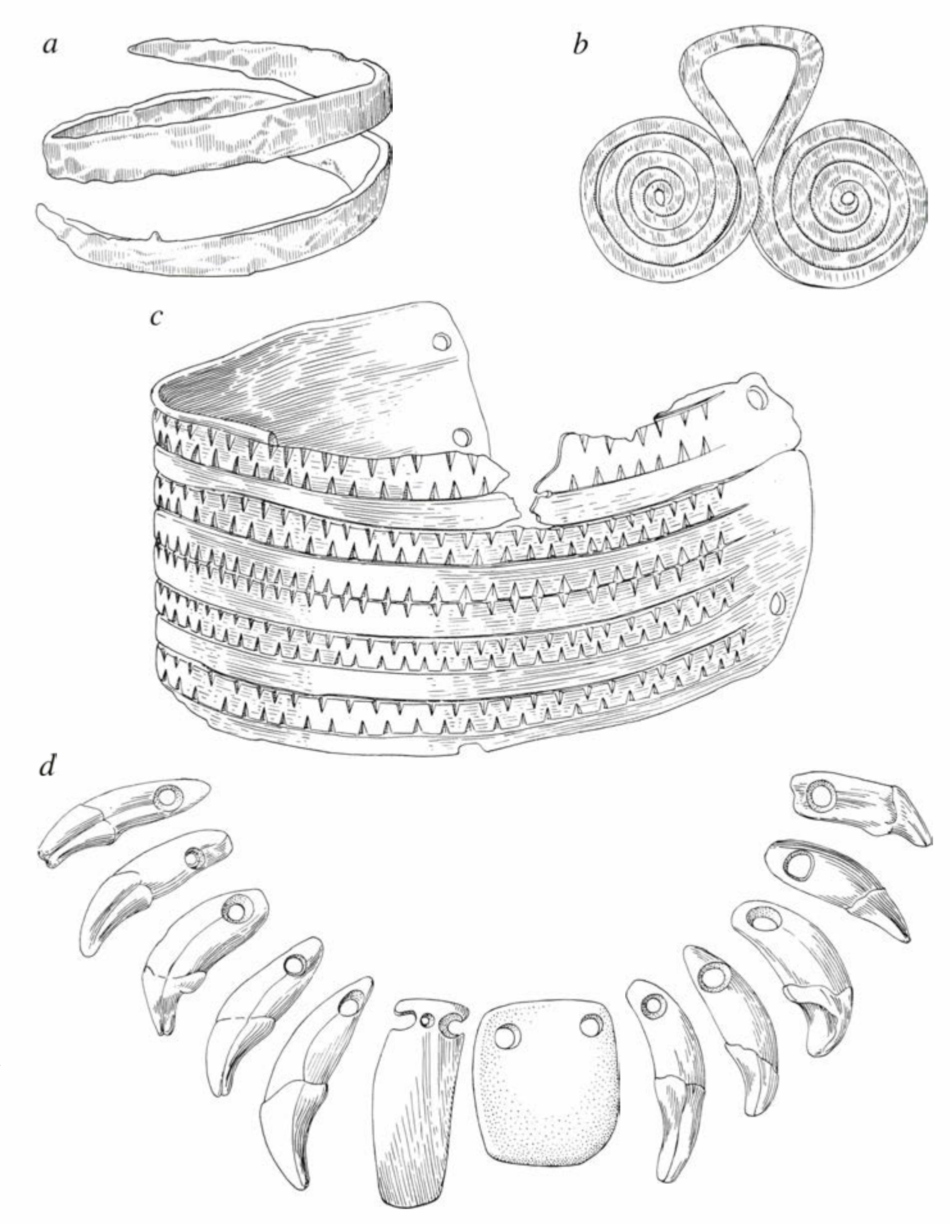
No se sabe a ciencia cierta si los Lengyel practicaban la deforestación para extender sus actividades agrícolas; aclarar tal cuestión precisa de análisis polínicos en varios niveles del suelo, perfiles que están aún sin evaluar en nuestros dos yacimientos. El trabajo en ese sentido podría verse finalizado en breve. Un posible indicio de deforestación es la presencia de huesos de águilas y halcones de los pantanos entre los restos de animales. El desarrollo de esas aves reclama zonas abiertas en parte de su hábitat natural.

Para determinar los recursos que los primeros neolíticos explotaban a lo largo de las diferentes etapas del año dependíamos de una serie de pruebas de carácter vario, incluyendo entre ellas los dientes de cerdo y ciervo y el grado de desarrollo de los mismos. También anotábamos otros indicios estacionales, como la presencia, entre los restos óseos, de ciertas aves acuáticas (indicativa de la caza de invierno y verano), de tortugas de embalse (que permanecen inactivas desde fines de otoño a inicios de la primavera) y de astas de ciervo (que se renuevan todos los años en febrero y marzo). No hallamos señal alguna de actividad invernal entre los desperdicios que correspondían a la cerámica lineal. Por su parte, la presencia de restos de tortugas indicaba que en los asentamientos había actividad durante los meses estivales. El contenido de los hoyos de desecho de los Lengyel demostraba, en cambio, que esos posteriores ocupantes vivían en el asentamiento todo el año.

El examen de dos "anillos de crecimiento" de los dientes de mamíferos resultó de especial valor. Los estratos de cemento dentario se depositan en bandas opacas, en verano, y en bandas traslúcidas en invierno. Seccionando un diente y determinando las características del último estrato depositado puede saberse, de forma aproximada, en qué época del año murió el animal. Nos enteramos así que la gente de Lengyel sacrificaba sus cerdos a mediados del invierno y cazaba ciervos y corzos a finales del invierno y en primavera.

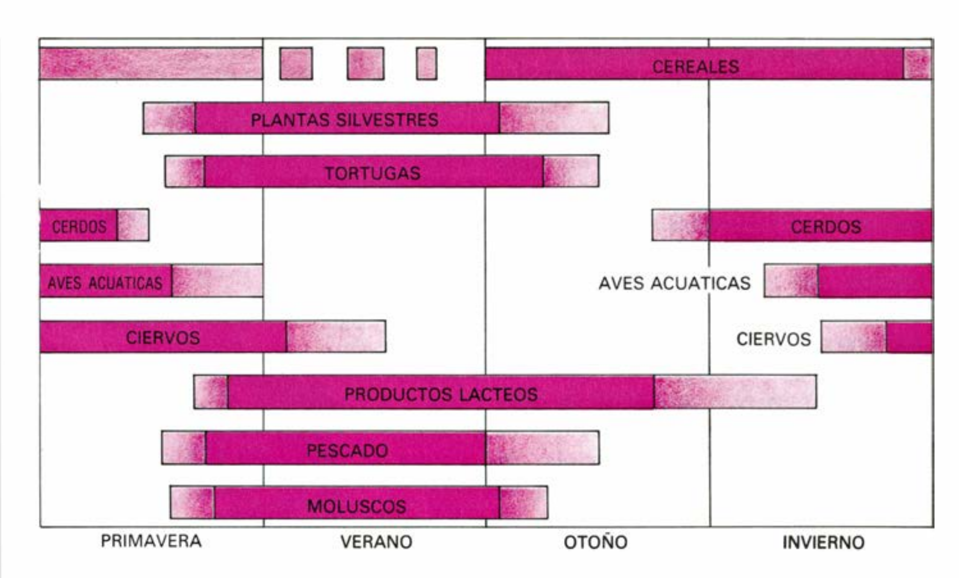
Además de investigar los yacimientos 3 y 4, reconocimos los alrededores en busca de otros asentamientos del Neolítico Inicial. Encontramos vestigios de otros seis asentamientos de la cultura de la cerámica lineal; todos mostraban rasgos similares a los ya vistos en Brześć Kujawski: no había rastro de una vivienda permanente y sus hoyos de desechos eran poco profundos. Se descubrieron otros tres asentamientos del grupo de Lengyel, todos a menos de una hora de camino de los yacimientos principales. Tampoco había en ellos pruebas de construcciones permanentes, y los pocos hoyos encontrados contenían depósitos de variedad más limitada que los de los asentamientos principales. Daban la impresión de tratarse de zonas donde se desarrolló alguna actividad especializada, quizás el cuidado de rebaños. Uno de ellos, el de Kuczyna, parecía corresponder a un aserradero, por el número exagerado de hachas de piedra y de asta de ciervo que aparecieron en sus depósitos.

En conjunto, la proximidad de los campamentos Lengyel al principal, y la ausencia de estructuras permanentes, permiten aventurar que los agricultores de Lengyel ocupaban un "territorio" bien definido, mientras que los campamentos de la cerámica lineal, más dispersos, ofrecen la impresión contraria. Ello, unido a la diferenciación funcional de los campamentos "satélites" de los Lengyel, indica que estos agricultores explotaban sus alrededores de un modo más organizado que sus predecesores de la cerámica lineal.



11. ORNAMENTOS TIPICOS de los enterramientos de Lengyel son los brazaletes, de cobre (*a*) y el pendiente, también de cobre, en forma de gafas (*b*). El brazalete de hueso con incisiones (*c*) se hacía de costillas animales (probablemente reses). El collar combina colgantes de coral fosilizado y pulimentado con dientes de perro o lobo perforados para ensartarlos. Los objetos de cobre, raros en el Neolítico Inicial, quizá se fabricaran con cobre acarreado de los campos mineros de los Cárpatos meridionales.

¿Qué conclusiones generales pueden extraerse del trabajo realizado en Brześć Kujawski? En el pasado, tanto la economía de subsistencia como la cultura material de las primeras comunidades neolíticas centroeuropeas solían caracterizarse como un conjunto homogéneo. Por ejemplo, siempre que se encontraba cerámica li-



12. CICLO ALIMENTARIO ANUAL de los pobladores Lengyel, comenzando a finales del invierno (*izquierda*). Se explotaba entonces un animal doméstico (el cerdo) y tres grupos salvajes (aves acuáticas migratorias, ciervos y corzos); desde finales de primavera hasta entrado el otoño, disponían de plantas silvestres comestibles, moluscos de agua dulce, tortugas y peces, además de la leche de vacas, ovejas y cabras. Con la recolección del trigo, en otoño, los Lengyel aprovechaban un importante recurso para pasar parte del invierno, hasta que los cerdos engordados en verano constituían de nuevo la principal fuente de alimento. Complementaban su dieta con aves acuáticas, ciervos y queso, producto almacenable.

neal se suponía automáticamente la presencia de casas alargadas, de un conjunto de hachas pulimentadas, picos, cereales, ganado vacuno, ovejas, cabras y cerdos. No es este el caso de los dos yacimientos polacos, ni el de otros de la cultura de la cerámica lineal de la llanura norteeuropea. No se han encontrado allí las casas alargadas de las tierras altas, las pruebas del cultivo de cereales escasean y tampoco abundan los huesos de cerdo. Es más, en las estancias del grupo de la cerámica lineal en los yacimientos 3 y 4 no se aprecian indicios de ocupación invernal. Lo contrario ocurre en los asentamientos de los Lengyel en los mismos yacimientos, y también en los asentamientos de la cerámica lineal en las tierras altas más meridionales.

Se desprende de estos datos la necesidad de reconsiderar la interpretación habitual del modo de vida de las primeras comunidades productoras de la llanura norteeuropea. A la vista de las omisiones que presenta la interpretación del tema como un conjunto homogéneo, y teniendo en cuenta la alta proporción de huesos de ganado vacuno hallados entre los restos de animales del grupo de la cerámica lineal, sugerimos que esta primera incursión neolítica al norte de las tierras altas no se tradujo en asentamientos agrícolas permanentes, sino en un pastoreo estacional (con fabricación de queso) en los bosques de las tierras bajas. La diversidad de la vegetación de las tierras bajas proporcionaría ricos pastos estivales para el ganado vacuno, probablemente mucho más ricos que los disponibles en los estrechos valles de las tierras altas.

Las condiciones favorables para el pastoreo estacional podrían haberse intensificado en el área forestal por la acción de los indígenas mesolíticos (preneolíticos) habitantes de la zona, en concreto, por la deliberada deforestación con ayuda del fuego. Prueba de esa práctica mesolítica sería la proporción, anormalmente elevada, de pólenes de avellano que muestran los perfiles polínicos correspondientes a un período datado hacia el 5800 a. de C., esto es, unos 500 años antes de la penetración por tierras altas del grupo de la cerámica lineal. El avellano es un árbol común entre la vegetación secundaria que invade un terreno aclarado.

Una prueba más de la importancia de los productos lácteos para estos primeros inmigrantes neolíticos es la presencia de coladores de barro, tanto en Brześć Kujawski como en otros yacimientos de la cerámica lineal de la llanura norteeuropea; estos objetos son raros, comparativamente, en los yacimientos de la cerámica lineal de las tierras altas, quizá porque los ocupantes de estas tierras poseían, además de cereales, numerosos cerdos, ovejas y cabras, mientras que la dieta de los ocupantes de las tierras bajas se limitaba a la recolección de plantas silvestres y a los productos lácteos. Otra explicación sería que los pobladores neolíticos de la cerámica lineal no habían desarrollado aún tolerancia a la lactosa (azúcar de la leche), característica de los habitantes modernos de Europa, por lo que habían de convertir la leche de vaca en queso (y quizás en yogur) y así consu-

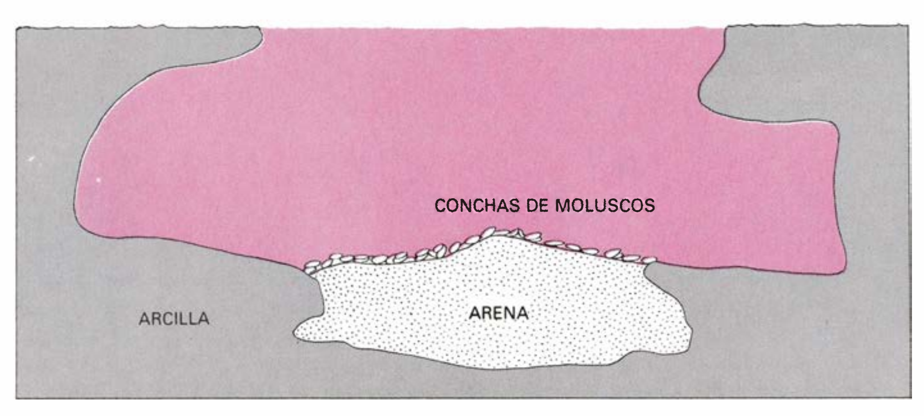
mir productos lácteos sin peligro de indigestión. Ante la acumulación de pruebas, parece probable que los pastores de la cerámica lineal vinieran al norte con su ganado vacuno en la estación cálida, para retornar a sus asentamientos permanentes de las tierras altas en invierno.

Los indicios dejados tras de sí por los agricultores del grupo de Lengyel que sucedieron en las tierras bajas a los de la cerámica lineal, apoyan la sospecha de que poseían la suficiente diversidad de recursos para consolidar una economía agrícola de rotación anual. Aunque cultivaban cereales, es dudoso que la cosecha de trigo bastara para alimentar a la comunidad hasta la siguiente cosecha y al mismo tiempo proporcionase semillas para la siembra de primavera. Ciertamente debían recolectar productos silvestres, como las avellanas y el quenodio (*Chenopodium*), recurso alimentario tanto en verde como en semilla, que se aprovechó profusamente por toda la Europa neolítica. En Polonia crece hasta una altura de 1,5 metros. El grupo de Lengyel quizá complementara su dieta con productos hortícolas, como los guisantes y las lentejas; ambos se encuentran en otros yacimientos neolíticos centroeuropeos, aunque no se han hallado pruebas de ellos en Brześć Kujawski.

Mayor importancia adquirieron, en la alimentación de los Lengyel, los animales domésticos, que estos pobladores agrupaban en rebaños, y la caza. Como afirman nuestros estudios de piezas dentarias, este grupo sacrificaba sus cerdos en invierno, tras el engorde de verano, y cazaba ciervos y corzos a finales del invierno y en primavera. La tortuga, peces y moluscos, conformarían parte de su dieta en primavera y verano y, como complemento, contarían con las plantas silvestres de los bosques adyacentes y la leche proporcionada por las vacas, ovejas y cabras. Tal combinación de alimentos silvestres y productos caseros reviste particular interés a finales del verano, cuando la cosecha y acopio de los cereales centraba la atención del grupo.

La organización y programación de este tipo de sistema de subsistencia fue, a todas luces, lo que impulsó el cambio en el modelo de distribución de los asentamientos con respecto al patrón disperso de la cerámica lineal. Ya hemos mencionado los campamentos "satélite" de Lengyel y su contribución al mantenimiento de un área residencial permanente; también habría probablemente otros campamentos locales (estaciones de caza y zonas de pastoreo sin cercar) que no han dejado, lógicamente, huella en el registro arqueológico. Los agricultores de Lengyel debían conocer a la perfección los bosques, campos y lagos adyacentes; hasta el último metro cuadrado contribuiría a la buena marcha del asentamiento central.

La larga continuidad de la ocupación del grupo Lengyel testifica el éxito alcanzado por esos agricultores neolíticos en la explotación del ecosistema forestal del entorno. Parece verosímil que el gradual deterioro del medio ambiente, tras siglos de ocupación humana, forzase el abandono de los asentamientos de Brześć Kujawski hacia el 3900 a. de C. No obstante, la producción de alimentos como modo de vida se había establecido ya de modo definitivo en la llanura norteeuropea. La revolución neolítica se extendía por entonces a lo largo de la costa del Báltico, por Alemania y Dinamarca, siendo sus protagonistas los pobladores neolíticos tardíos, como los de la cultura de los jarros con cuello de embudo. Los grupos pioneros de la cerámica lineal y de Lengyel desaparecieron, pero la agricultura había logrado establecerse en la Europa templada.



13. HOYO DEL YACIMIENTO NUMERO 3, próximo a la orilla del antiguo lago, que se nos muestra en una sección transversal simplificada. La presencia de arenas intrusivas y una capa de conchas de moluscos sugiere que el hoyo sirvió de almacén y se usó como estanque donde criar moluscos, tortugas y peces.

Señales químicas de agregación en las amebas

Dos hongos mucosos unicelulares coexisten en un mismo suelo sin perder su identidad. Emiten distintas sustancias químicas y reaccionan frente a otras, asimismo diferentes

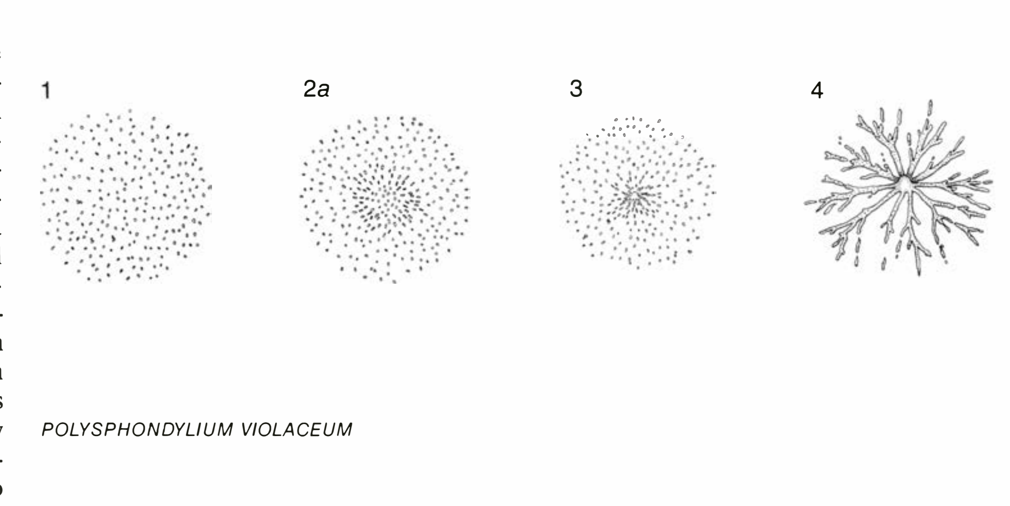
John Tyler Bonner

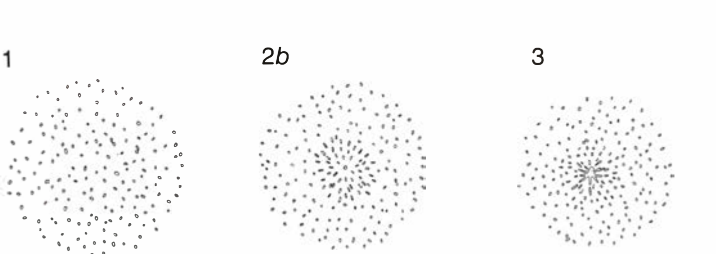
Detengámonos ante un suelo medianamente fértil; abundan allí dos tipos de minúsculos organismos unicelulares: bacterias y amebas. En un centímetro cúbico de suelo puede haber millones de bacterias y miles de amebas. Las bacterias tienden a agruparse en torno a las raíces muertas o junto a restos de insectos, de los que sacan su alimento; las amebas, por su parte, se comen a las bacterias. El tamaño de las amebas es comparable con el de los glóbulos blancos de la sangre; son, pues, unas 10 veces más largas que una célula bacteriana. Se arrastran por el suelo, engullendo las bacterias que encuentran a su paso y, cada pocas horas, se dividen en dos células hijas.

Entre las clases más comunes de amebas del suelo, ocupan un lugar destacado las sociales, conocidas también por hongos mucosos celulares. Se alimentan célula a célula, cual amebas individuales; pero, una vez que han acabado con todas las bacterias de una cierta área, se desencadena en ellas el inicio de su fase social. Las células cesan de vagar solitarias y sin rumbo y comienzan a formar un reguero hacia un punto central de reunión. Se acumulan en él desde unos cuantos hasta unos 100.000 organismos unicelulares y crean una masa única, en forma de salchicha, que da lugar a un organismo pluricelular.

Las células empiezan entonces a diferenciarse. En el extremo frontal de la masa aparece un ápice en forma de pezón, en cuyo interior se va configurando un tallo: un cilindro de celulosa compuesto de células rígidas hinchadas que acabarán muriendo. A medida que las células se dirigen corriente arriba, se va enderezando el tallo y, fuera de él, se apilan a su salida. Cuando ha cesado todo el movimiento, el amasijo celular se ha convertido en un cuerpo fructífero: un tallo, simple o ramificado, formado por células muertas encerradas en una vaina de celulosa y rematado por una pequeña bola de esporas. Las esporas son amebas encapsuladas individualmente que se dispersan con agua o al entrar en contacto con cualquier insecto o gusano que pase por el lugar. Si la espora aterriza en un ambiente con la temperatura y humedad adecuadas, se rompe y se abre, liberando una sola ameba, que vuelve a empezar la fase errante del ciclo biológico del hongo mucoso.

Me referiré en este artículo a un aspecto particular de la conducta de las amebas sociales. Con frecuencia coexisten dos o más especies en la misma porción de suelo. Al agregarse, lo hacen por separado: las células de espe-





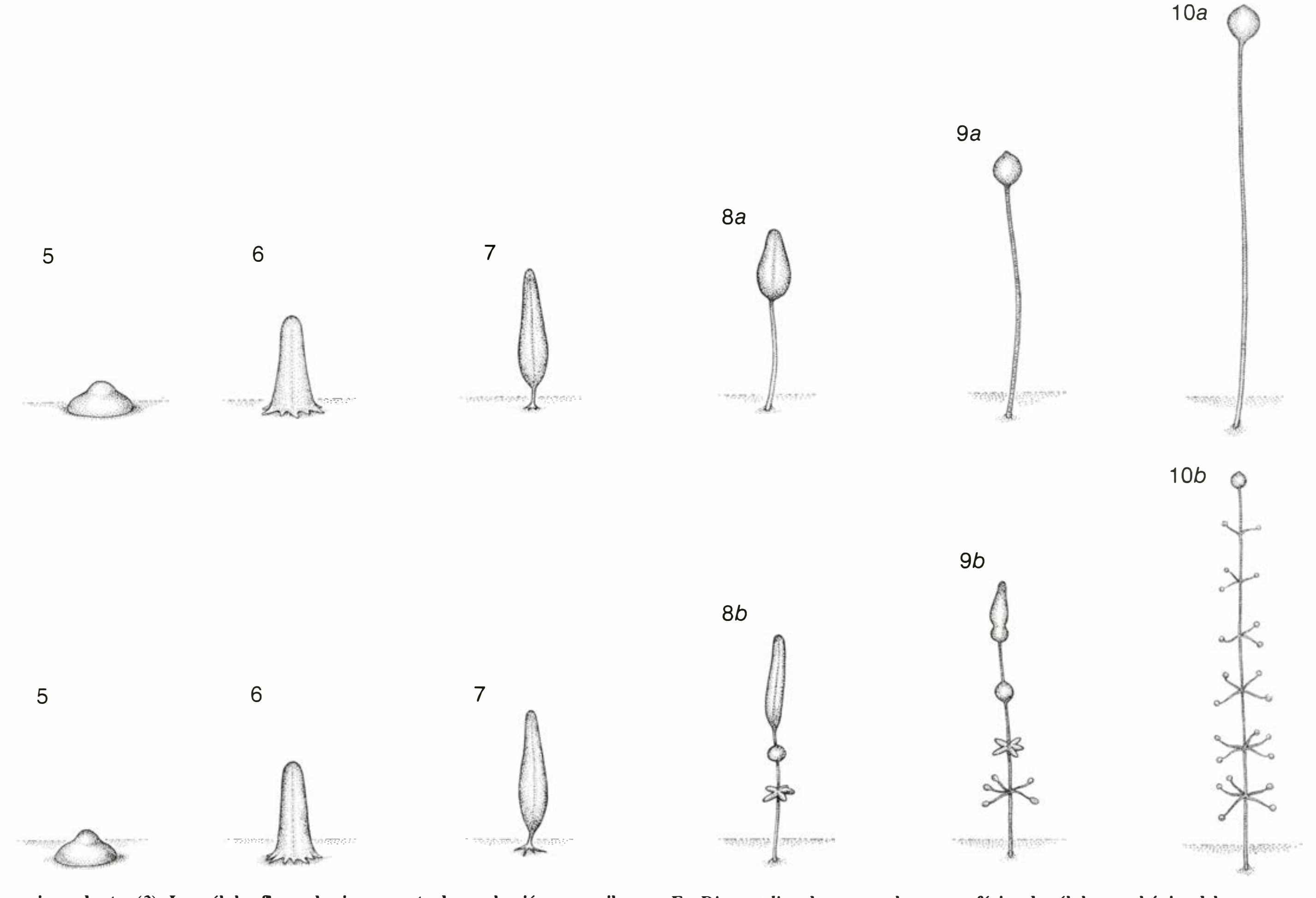
1. CICLO BIOLOGICO Y MORFOLOGIA de dos hongos mucosos unicelulares, *Dictyostelium mucoroides* y *Polysphondylium violaceum*. Son notablemente similares. Ambas especies se alimentan una a una, como amebas (*1*), hasta que la falta de alimento (bacterias) hace que se dispare en un grupo de células (*2a*) o en una sola célula (*2b*), la secreción de una sustancia química, llamada acrasina, que atrae a las amebas

cies distintas se orientan hacia centros de recolección diferentes. Las corrientes de células que se dirigen a un centro atraviesan e ignoran las corrientes que se dirigen a otros centros, de tal manera que cada especie se reúne en su propio centro para formar allí un cuerpo fructífero característico. ¿Cómo reconocen las amebas a los demás miembros de su especie? A lo largo de 40 años se han ido recogiendo informaciones fragmentarias, que, reunidas, nos permiten hoy explicar, a nivel molecular, la segregación de las dos especies.

Del fenómeno de segregación se ocuparon ya, en 1940, Kenneth B. Raper y Charles Thom, adscritos por entonces al Departamento de Agricultura de los Estados Unidos. Obtenían cultivos de la bacteria *Escherichia coli* e inoculaban las placas con dos especies de amebas sociales: *Dictyostelium discoideum* y *Polysphondylium violaceum*. Las amebas se multiplicaban y difundían a partir del punto de inoculación; en un momento dado, las dos poblaciones aparecían totalmente mezcladas. Tras agotar las bacterias disponibles, las amebas empezaban a agregarse. Raper y Thom lo describieron así: "Siempre se agregaban en centros distintos; las corrientes convergentes de las dos formas generalmente se solapaban... Era obvio que los estímulos de una y otra especie diferían cualitativamente". Descubrieron por qué tenía que ser así. Si las dos especies se hallaban juntas en el mismo suelo, debían disponer de algún mecanismo que evitara su coincidencia en un mismo agregado, asegurando su permanencia como especies distintas. Raper y Thom no pudieron saber en qué diferían los dos estímulos; se desconocía entonces su naturaleza.

A principios de siglo, los científicos que estudiaban los hongos mucosos consideraban que la reunión de las amebas debía responder a algún tipo de quimiotaxis; en los años 1940, sin embargo, se veía con suspicacia que un proceso de desarrollo tuviera por base la atracción química; la existencia de algún tipo de "guía por contacto" gozaba de mayor favor. En 1942, Ernest H. Runyon, del Agnes Scott College, demostró que el contacto entre las células no podía llevar a cabo la agregación. En una serie de experimentos realizados a lo largo de los años siguientes yo mismo puse de manifiesto que la masa central debía emitir algún agente atractor difusible. Podía haber sido el calor, pero una serie de indicaciones dejaban bien claro que tenía que tratarse de una sustancia química. La bauticé acrasina. (Acrasia es el nombre de una bruja del libro *Faerie Queene*, de Edmund Spenser, que atraía a los hombres y los transformaba en bestias, y *Dictyostelium* es un miembro del orden Acrasiales.)

Estudios realizados durante años en numerosos laboratorios, incluido el nuestro de la Universidad de Princeton, han puesto de manifiesto muchos detalles del sistema acrasina. La inanición es el estímulo que hace pasar del estado de alimentación al estado social. Desencadena una serie de acontecimientos químicos en las células, que pronto empiezan a segregar una acrasina. Al mismo tiempo, las amebas sintetizan y exponen en su superficie un gran número de moléculas de proteína que actúan de receptores específicos para la acrasina. Las células segregan



circundantes (*3*). Las células fluyen hacia un punto de recolección y se apilan para formar un organismo pluricelular (*4*). Sobre el amasijo celular se forma un ápice (*5*). Algunas células constituirán un tallo interno rígido (*6, 7*). Las restantes se sitúan en la parte superior del tallo para convertirse en esporas. En *Dictyostelium* hay una sola masa esférica de células en el ápice del cuerpo fructífero (*8a, 9a, 10a*). En *Polysphondylium* hay verticilos de pequeños cuerpos fructíferos, y una masa apical de esporas sobre el tallo central (*8b, 9b, 10b*). Las esporas se dispersan y se rompen, liberando amebas al abrirse.

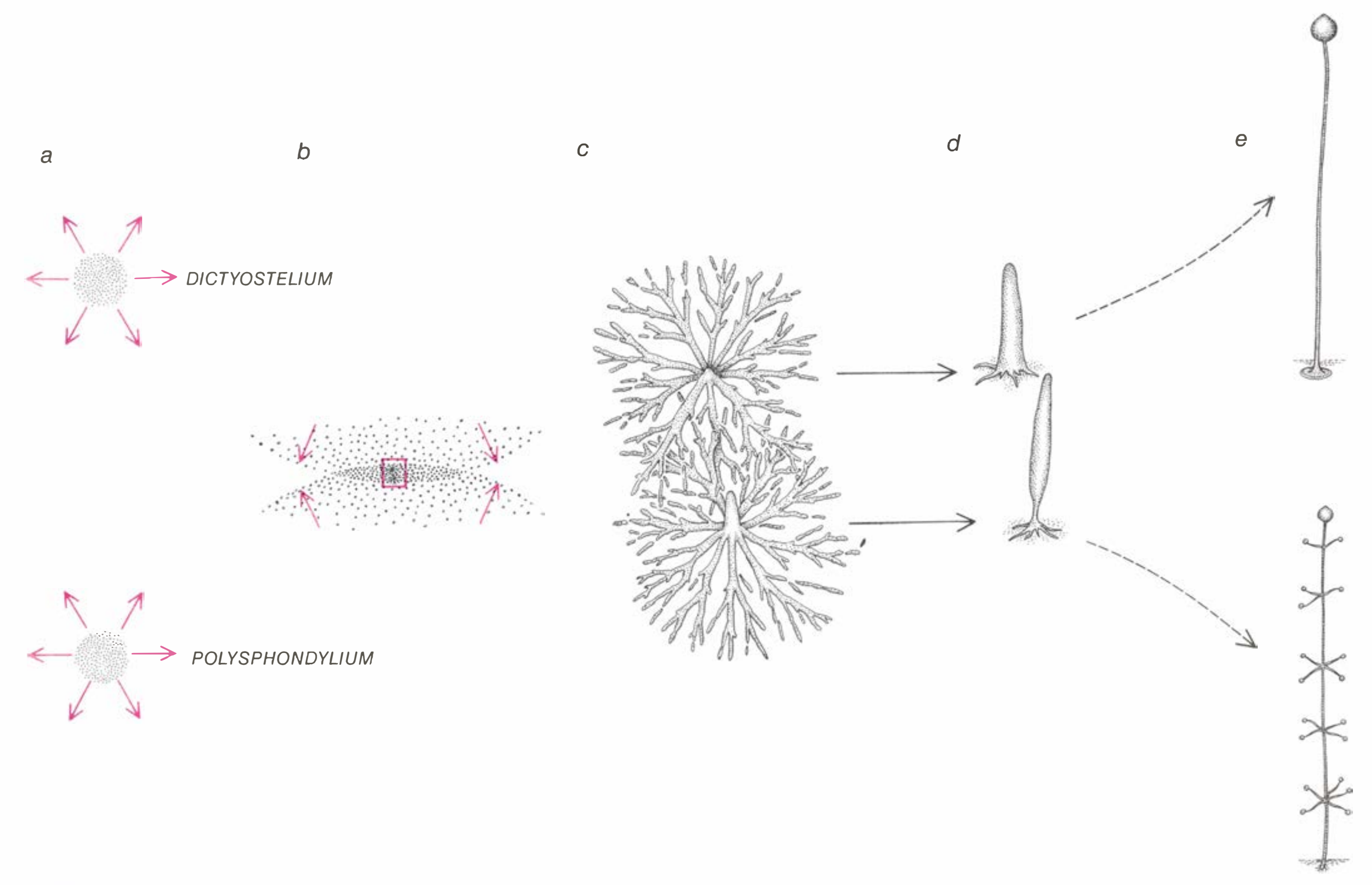
también una acrasinasa (un enzima que inactiva la acrasina) y un inhibidor de la acrasinasa. La acrasina, la acrasinasa y el inhibidor se combinan para formar un complejo sistema de control que modula la agregación a medida que se va produciendo. Por ejemplo, la acrasinasa mantiene el nivel de acrasina externa generado por una determinada ameba lo suficientemente bajo para que la célula sea sensible a la acrasina emitida por otras células.

La agregación se inicia de forma distinta en cada especie, como observó hace algunos años Brian H. Shaffer, cuando estaba trabajando en la Universidad de Cambridge. De repente una célula de la población de *Polysphondylium* se redondea y empieza a segregar acrasina. Las células vecinas se alinean para unirse a ella, orientadas según el gradiente de concentración de la acrasina segregada. En *Dictyostelium*, por el contrario, un pequeño grupo de células (Shaffer lo denominó una nube) es el que parece iniciar la agregación. Estas células, en primer lugar, permanecen relativamente inmóviles y luego se agregan, atrayendo el grupo a las células vecinas. En ambas especies, las amebas en agregación suelen pasar al interior en oleadas. Parece que existe una buena razón que lo justifica: para establecer un gradiente total de acrasina que sea eficaz y que oriente a las células distantes, se requiere una masa central de buen tamaño. Las oleadas pueden orientar a las células antes de que ese gradiente se establezca. Si una ameba libera una bocanada de acrasina y ésta se difunde hasta las células vecinas y toca a cada una de las células, primero por un lado, llena mayor número de receptores de la célula situados a ese lado que en los otros, lo que sirve de orientación a la célula. A su vez, la ameba que responde emite otra bocanada de acrasina que atrae a las células que vienen detrás de ella.

Al determinarse el origen quimiotáctico del mecanismo de agregación, sostuve que no importaba mucho conocer la naturaleza química de las acrasinas. Lo recuerdo ahora con rubor, pues la realidad es que en cuanto se descubre la estructura química de una sustancia clave de la fisiología o del desarrollo, uno salta de la edad media a la modernidad. Ciertamente también era ese el caso de la acrasina de *Dictyostelium*. De cualquier forma, pronto nos unimos a otros laboratorios en la tarea de caracterizar su atractor; lo que nos ocupó desde 1947 hasta 1967.

La principal dificultad que se nos presentaba era la de desarrollar una valoración biológica justa, es decir, un método para comprobar el efecto biológico de sustancias progresivamente más purificadas. Theo M. Konijn (hoy en la Universidad de Leiden) y mi grupo creamos uno. Dirigimos la primera parte de nuestro trabajo a identificar un factor bacteriano que parecía atraer a las amebas. Konijn vino a Princeton, y muy poco después de su llegada, David S. Barkley, a la sazón estudiante de licenciatura, tuvo la inspiración: ¿Por qué no probar el monofosfato de adenosina cíclico (AMP cíclico), un nucleótido que había descubierto Earl W. Sutherland, Jr., de la Facultad de Medicina de la Universidad Vanderbilt, y que, como había demostrado, actuaba en las células de mamífero como un "segundo mensajero", mediando entre las hormonas y las funciones celulares?

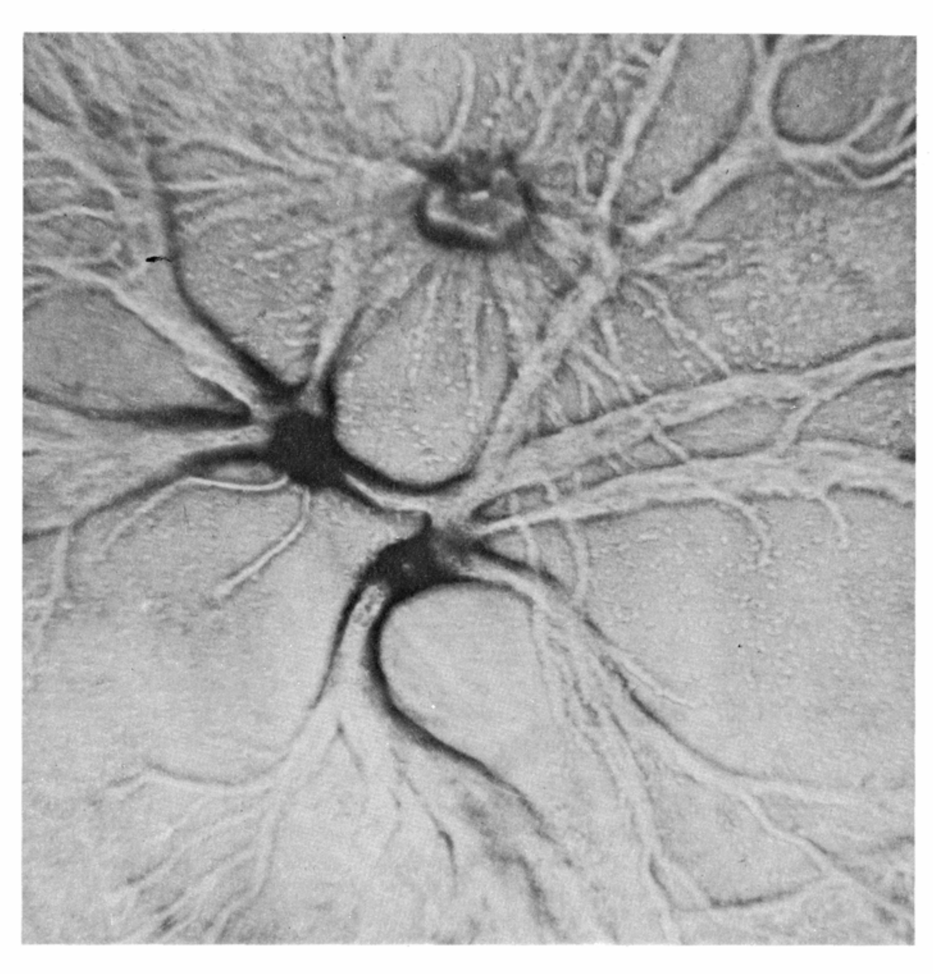
Funcionó, con notable potencia, en nuestros ensayos de quimiotaxis. Se evidenció que las células de *Dictyostelium* no sólo respondían a una diminuta cantidad de AMP cíclico, sino que,



2. DOS ESPECIES DE HONGOS MUCOSOS se agregan por separado, tal como Kenneth B. Raper y Charles Thom demostraron en 1940 con este experimento. Se inoculan en un cultivo de la bacteria *Escherichia coli* células aisladas de *Dictyostelium discoideum* (*arriba*) y de *Polysphondylium* (*abajo*). Las amebas se desplazan errantes desde el punto de inoculación (*a*), se mezclan y empiezan a agregarse (*b*). Cuando se examinan los agregados al microscopio, se aprecia que se mantienen separados. Las corrientes de amebas de las dos especies avanzan hacia dos centros de recogida diferentes y forman organismos independientes (*d*) que se desarrollan para formar cuerpos fructíferos característicos de cada especie (*e*). Mantienen así su identidad específica.

además, lo producían. (Hoy sabemos que la práctica totalidad de las células fabrican AMP cíclico, y que la excepción es no sintetizarlo.) En particular, observamos que, durante la agregación, las amebas segregaban el nucleótido en grandes cantidades (unas cien veces más) y que multiplicaban asimismo su sensibilidad al AMP cíclico con respecto a lo que ocurría en otras fases.

El descubrimiento de que el AMP cíclico era la acrasina natural de *Dictyostelium* abrió nuevas líneas de investigación. Laboratorios de todo el mundo han estudiado de qué manera el nucleótido orienta las células y han analizado la actividad de la adenilciclasa, el enzima que cataliza la síntesis de AMP cíclico, y de la fosfodiesterasa, la acrasinasa de esta acrasina. Han examinado también los receptores de la superficie celular para el AMP cíclico y las interacciones de todos los elementos del sistema. En 15 años se ha producido un gran avance en la comprensión del control de algunos aspectos del desarrollo de *Dictyostelium*; tales conocimientos no sólo conciernen al desarrollo de los hongos mucosos celulares, sino también al desarrollo en general.



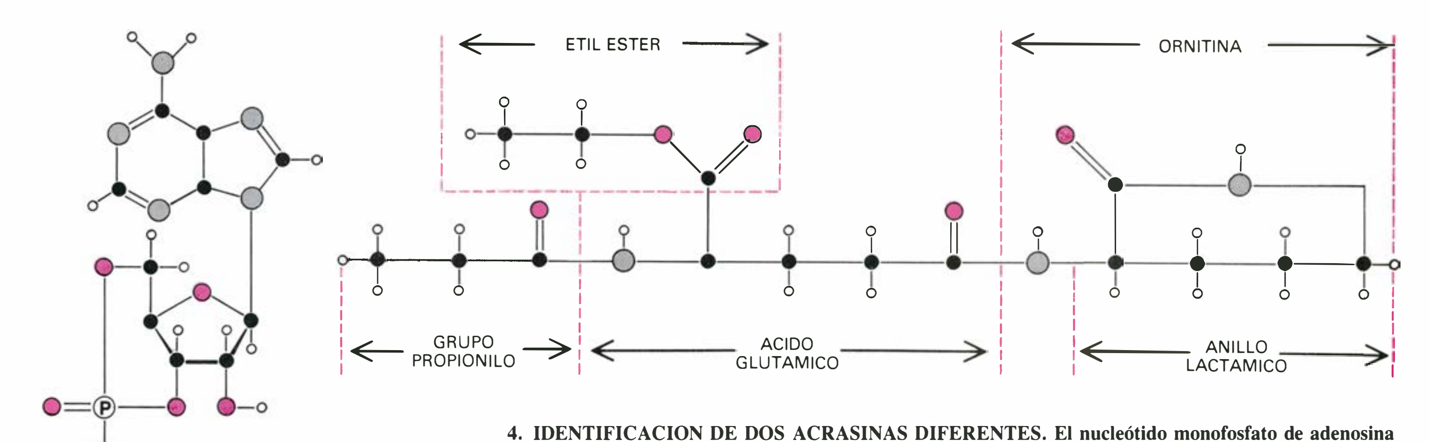
3. AGREGACION INDEPENDIENTE de dos especies, registrada en esta microfotografía tomada por Raper y Thom, donde las corrientes de amebas en agregación aparecen ampliadas unas 45 veces. Se forma un pequeño agregado de *D. discoideum* (*a*) y un agregado mayor, con dos centros oscuros, de *P. violaceum*. Las corrientes de amebas que fluyen hacia ellos se superponen y las células de cada especie parecen ignorar las de la otra especie al dirigirse hacia su propio centro y formar el tallo esporógeno.

Tan pronto supimos que el AMP cíclico atraía a las células de *Dictyostelium* lo ensayamos sobre *Polysphondylium*. No surtió efecto, como Raper y Thom hubieran predicho (y como Shaffer postulara luego). Las dos especies tienen acrasinas diferentes. En el correspondiente esfuerzo por identificar la acrasina de *Polysphondylium* no repetiríamos la experiencia con *Dictyostelium*. A ninguno se nos ocurrió una idea genial que nos llevara al éxito; ni siquiera fuimos capaces de encontrar una buena fuente de atractor sin purificar. (Los productos bacterianos y la misma orina humana habían sido óptimas fuentes de AMP cíclico.) Ensayamos múltiples productos que generalmente son ricos en hormonas y otras sustancias activas, como leche, peptonas (mezclas de péptidos), extractos de levadura y otros muchos más, pero sin resultados. No quedaba otro recurso que extraer el material de las amebas en agregación. (Moraleja: no hay que confiar en que la suerte nos visite dos veces.)

Como primer paso, hubo que buscar la mejor forma de extraer la acrasina de un agregado de *Polysphondylium*. Por los trabajos con *Dictyostelium* sabíamos que la acrasinasa segregada junto con la acrasina tendía a destruir el atractor en cuanto se segregaba. Al fin pusimos a punto un método diseñado por David W. Francis, becario postdoctoral en aquellas fechas en Princeton y hoy profesor de la Universidad de Delaware. Consistía en verter alcohol etílico al 40 por ciento directamente sobre las placas de Petri que contenían células a mitad del proceso de agregación, lavar las placas con más alcohol y recoger luego el líquido de lavado en un vaso de precipitado. Se eliminaba el alcohol por ebullición y se recogía por centrifugación el precipitado. En solución quedaba un extracto crudo de acrasina, que permanecía intacto porque el alcohol desnaturalizaba la acrasinasa.

Lo más notable del caso era que las amebas no parecían inmutarse por esa zambullida etílica; proseguían su desarrollo después de eliminar el alcohol. Para recoger la máxima cantidad posible de extracto crudo de la acrasina de *Polysphondylium*, hemos perfilado, con los años, un método de rutina: se preparan 150 grandes placas de Petri con amebas en agregación que cubren cuan ancho es el laboratorio. Tiene el inconveniente de que obliga a suspender los demás experimentos, por falta de espacio, pero el laboratorio se llena de un aroma bastante agradable. Nos pareció al principio que el olor provenía de la acrasina, pero era alcohol lo que aspirábamos.

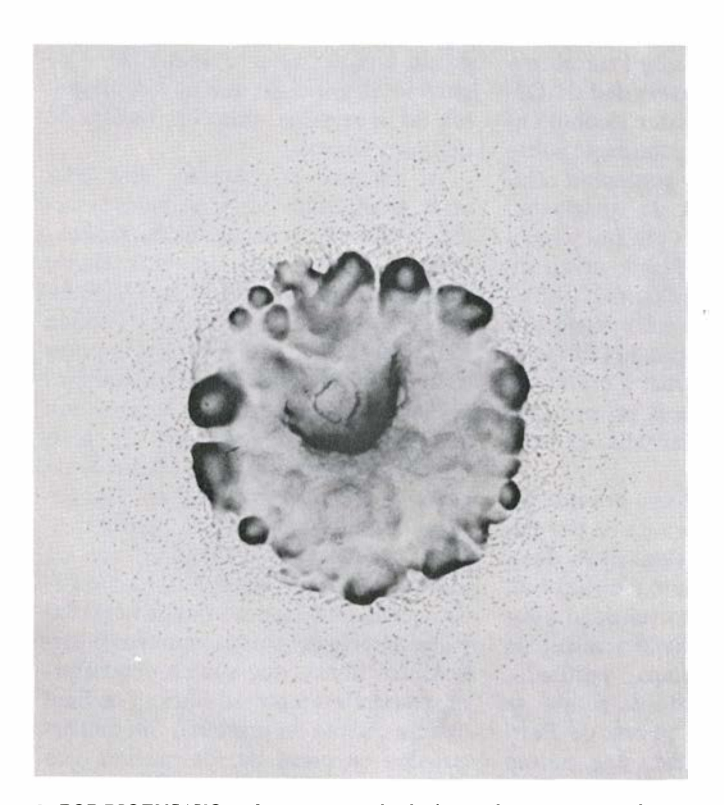
Al iniciarse el proyecto, diez años atrás, buscábamos las principales características de la acrasina de *Polysphondylium*. Trabajaban en ello Bernd Wurster, actualmente en la Universidad de Constanza, Alemania, y Pauline Pan, ahora en la Pennwalt Corporation. Tras dos años largos de trabajo, podíamos mantener que la acrasina era una molécula pequeña, con un peso molecular inferior a 1500 (calculado de acuerdo con su movilidad en una columna de filtración en gel); que era estable cuando se la calentaba; que no parecía ser de naturaleza iónica (es decir, que la molécula no estaba cargada) y que carecía de grupos amínicos o carboxílicos libres, que suelen encontrarse, respectivamente, al inicio y al final de una cadena de proteína. Sin embargo, dos enzimas, de los muchos que probamos, una proteasa y una peptidasa, inactivaban la acrasina; ello sugería que se trataba de algún tipo de péptido (un fragmento corto de una cadena de

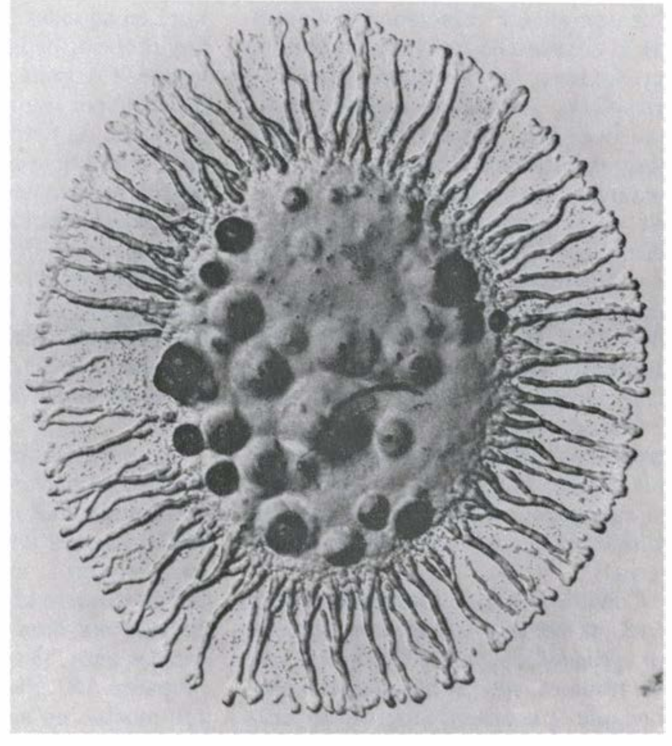


4. IDENTIFICACION DE DOS ACRASINAS DIFERENTES. El nucleótido monofosfato de adenosina cíclico, o AMP cíclico (*a la izquierda*), es el atractor natural de las especies de *Dictyostelium*. La acrasina de *Polysphondylium violaceum*, purificada y caracterizada recientemente por el autor y sus colegas, es un dipéptido, bautizado con el nombre de glorina (*a la derecha*) por contracción de sus principales componentes: los aminoácidos ácido glutámico y ornitina. Un grupo amínico (NH_2) y un grupo carboxílico (COOH) del ácido glutámico están bloqueados, respectivamente, por un grupo propionilo y un etil éster. Un grupo amínico de la molécula de ornitina está bloqueado por un anillo lactámico. Tanto el AMP cíclico como la glorina son moléculas de pequeño tamaño, con pesos moleculares de 324 y 327, respectivamente.

proteína). Desgraciadamente, el hallazgo no fue concluyente, porque las preparaciones enzimáticas eran impuras y podía intervenir solapadamente algún producto secundario. Se trataba, sin embargo, de enzimas neutros, que atacan polipéptidos sin carga, lo que estaba en buen acuerdo con la observación de que la molécula de acrasina parecía no hallarse cargada. Concluimos que la acrasina de *Polysphondylium* podía ser un péptido en el que no había ningún grupo amínico libre, con algunos indicios de que llevaba incorporado lo que se denomina un enlace de tipo éster.

El análisis posterior resultaba difícil. A pesar de nuestros esfuerzos, los extractos más purificados seguían siendo muy impuros. Llegados a ese punto, nuestro colega Osamu Shimomura, actualmente en el Laboratorio de Biología Marina de Woods Hole, Massachusetts, se interesó por el tema, y se hizo cargo de toda la parte química, purificaciones incluidas. Para suministrarle cantidad suficiente de material de trabajo hubo que aumentar nuestro número de recogidas. Bajo la dirección de Hannah B. Suthers iniciamos una "fábrica de acrasina" y dedicábamos dos días a la semana a recoger el contenido de las 150 grandes placas de Petri con acrasina. Se mantuvo el ritmo durante tres años, con un total de 96 recogidas. Durante los dos primeros años seguimos experimentando, perfeccionando el método de purificación. El tercer año nos dedicamos a hacer un esfuerzo final para acumular acrasina cruda; las 40 primeras recolecciones nos proporcionaron cuatro gramos iniciales de un





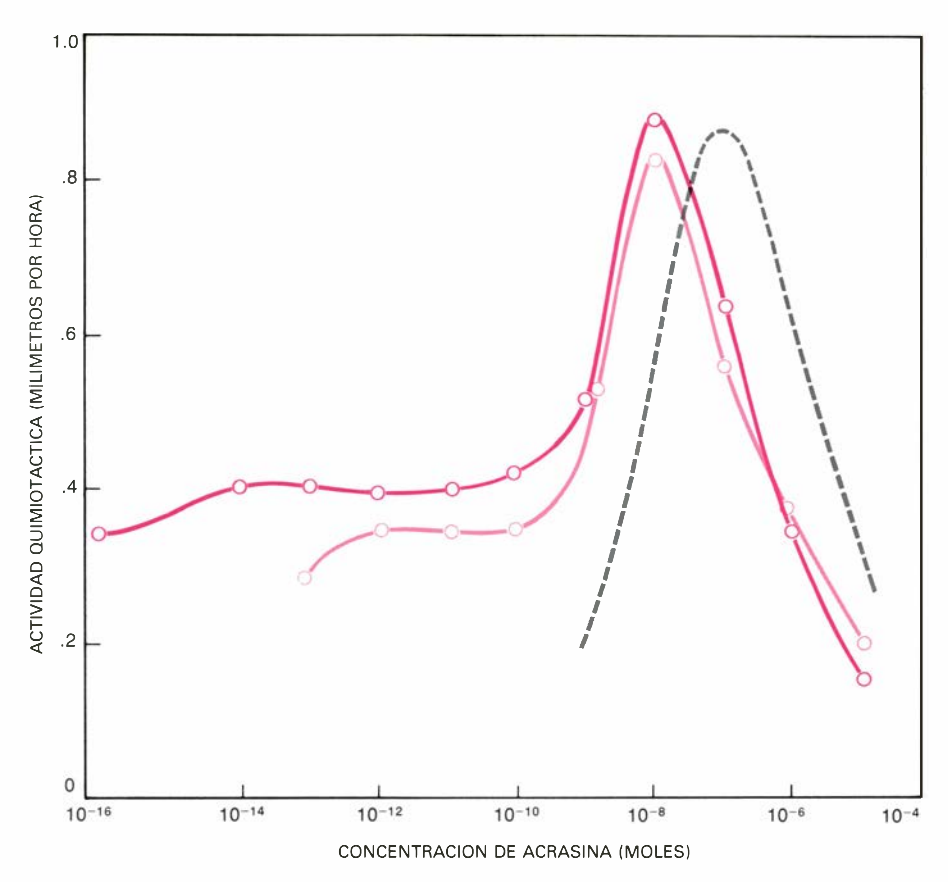
5. POR BIOENSAYO se demuestra que la glorina es el atractor natural para *P. violaceum*. Cuando se coloca sobre un bloque de agar un gran acúmulo de amebas (*a la izquierda*), todas las células, salvo unas pocas, permanecen en el grupo. Si se coloca el acúmulo de amebas sobre un bloque de agar impregnado con glorina sintética (*a la derecha*), las amebas se desparraman inmediatamente hacia fuera, atraídas por la glorina. (La glorina que queda directamente debajo del grupo desaparece en seguida, degradada por una acrasinasa secretada por las células.) Queda por identificar la acrasinada de *P. violaceum*.

material muy impuro. Shimomura los sometió a una serie de nueve pasos de purificación que comprendían tres clases de columnas de filtración en gel y numerosos sistemas de solventes. Al final obtuvimos 92 microgramos (millonésimas de gramo) de acrasina, con un grado de pureza de alrededor del 98 por ciento.

Los 92 microgramos de acrasina se dividieron en lotes de unos 10 microgramos, que se enviaron para hacer con ellos diferentes clases de análisis. De acuerdo con su espectro al infrarojo, parecía verosímil que tuviese un grupo éster. El análisis de los aminoácidos confirmó el hallazgo anterior de que la molécula contenía dos aminoácidos, ácido glutámico y ornitina (que no es uno de los componentes habituales de las proteínas), en cantidades aproximadamente iguales. Más interés proporcionaron los resultados obtenidos por Catherine E. Costello con el espectrómetro de masas del Instituto de Tecnología de Massachusetts. Utilizando las últimas técnicas (ionización de desorción de campo y bombardeo con átomos rápidos, a alta y a baja resolución) determinó la fórmula estructural completa y el peso molecular. Conviene no perder de vista que esos datos se obtuvieron a partir de una muestra de sólo unos 10 microgramos de material.

La acrasina de *Polysphondylium* resultó ser un dipéptido; una molécula de ácido glutámico con dos grupos laterales: un grupo propionilo, que bloquea el extremo amínico del aminoácido, y un ester etílico, que bloquea el grupo carboxílico libre. El ácido glutámico está unido a la ornitina mediante un enlace peptídico, pero la ornitina forma un anillo lactámico que bloquea de nuevo el grupo amínico libre. El peso molecular de la acrasina es de 327. Puesto que consta de ácido glutámico y ornitina, cuyas abreviaturas son *Glu* y *Orn*, decidimos llamarla glorina.

Para confirmar que la glorina era la acrasina natural de *Polysphondylium*, necesitábamos sintetizar el dipéptido y probar luego la versión artificial en nuestro ensayo de quimiotaxis con amebas de *Polysphondylium*. Enviamos la fórmula de la glorina a los Peninsular Laboratories, Inc. y, un mes más tarde nos entregaron 50 miligramos de glorina sintética. (Calculamos que hubiéramos tardado más de 500 años en acumular tal cantidad de material por extracción y purificación en nuestro labortorio.) En el MIT, Costello comprobó inmediatamente que sus propiedades manifestadas en el espectrógrafo de masas eran idénticas a las de la glorina natural purificada. Con no poca ansiedad, emprendimos el bioensayo en Princeton.



6. ACTIVIDAD QUIMIOTACTICA de la glorina sintetizada en el laboratorio (*color oscuro*), comparada con la actividad del atractor natural (*color claro*) purificado a partir de células en agregación. Las dos curvas son esencialmente idénticas, confirmando así los análisis químicos practicados a la glorina natural. Una curva similar (*gris*) muestra el efecto del AMP cíclico sobre células en agregación de *D. discoideum*.

El resultado fue inequívoco. La glorina sintética poseía una actividad biológica intensa; cuando se colocaba en el agar que rodeaba a una masa de amebas sometidas a ensayo, éstas se dispersaban espectacularmente [*véase la figura 5*]. En una prueba cuantitativa comparamos la capacidad de la glorina sintética para atraer las amebas con la capacidad de la glorina natural purificada (los restos de nuestras provisiones): ambas se mostraban esencialmente idénticas en cuanto a actividad se refería. Al comparar las curvas de la glorina con los datos que mostraban el efecto del AMP cíclico sobre la agregación de las amebas de *Dictyostelium*, se vio claramente que la glorina era eficaz incluso a concentraciones inferiores a las del AMP cíclico.

Dejemos aquí la emoción de la caza. ¿Qué podíamos hacer ahora con la glorina? Por una parte, aclarar todos los detalles del sistema de agregación de la glorina, incluida la identificación de sus receptores, de la acrasinasa y del inhibidor de la acrasinasa. También nos planteamos esta otra pregunta: ¿En qué medida intervenía la glorina en la diferenciación y en qué medida lo hacía el AMP cíclico? La razón de ese interés radicaba en que se sabía, desde nuestros primeros experimentos con Konijn, que en *Polysphondylium* había AMP cíclico. Posteriormente, en nuestro laboratorio y en otros, se fueron acumulando pruebas de que el AMP cíclico estaba implicado en la diferenciación de *Dictyostelium*, tanto de las células del tallo como de las esporas, y que lo mismo ocurría en *Polysphondylium*. Se habían realizado incluso valiosos trabajos con *Dictyostelium* que revelaban una implicación directa del AMP cíclico en el control de la expresión de algunos genes del desarrollo. Investigaciones recientes, acometidas en laboratorios de Holanda y de los Estados Unidos, han aportado pruebas convincentes de que la quimiotaxis continúa operando en la masa de células durante los últimos estadios del desarrollo: un gradiente de AMP cíclico orienta las células, guiándolas de este modo en la dirección del ápice que avanza. Durante esos últimos estadios del desarrollo de *Dictyostelium*, el AMP cíclico está, por tanto, implicado en dos funciones, quimiotaxis y diferenciación, al mismo tiempo y en el mismo lugar.

Pretendemos resolver si en *Polysphondylium* existe o no una separación

de funciones, esto es, si la quimiotaxis la controla la glorina y, la diferenciación, el AMP cíclico. La relación de los dos sistemas de señalización quizá no sea sencilla ni independiente; cabe una mutua dependencia o que la glorina esté más directamente implicada en la diferenciación celular de lo que se supone. Nos hallamos ante una oportunidad única para comparar dos sistemas paralelos de desarrollo que tienen distintos sistemas de señales. Entre otras cosas, proporcionaría un camino para separar, aislar e identificar los controles de los diferentes acontecimientos del desarrollo.

La cuestión principal estriba, quizás, en que *Dictyostelium* y *Polysphondylium* viven en el suelo, mezclados, tienen el mismo aspecto y, en su desarrollo, atraviesan fases de notable parecido. Pero poseen sistemas de señales totalmente diferentes; el sistema de uno se basa en un nucleótido y, el del otro, en un péptido. En la única situación donde encontramos algo igual es en el cuerpo de los animales, humanos incluidos; aquí el AMP cíclico actúa en muchas reacciones de señalización, tanto hormonales como neuronales, y un cierto número de péptidos funcionan como neurotransmisores del sistema nervioso. La presencia de sistemas rivales de señales que se apoyan en pequeños nucleótidos y en péptidos es un fenómeno biológico muy generalizado, no una mera curiosidad de las amebas.

La diferencia tajante entre estos dos sistemas de señales plantea algunas cuestiones evolutivas. Indiqué antes que los experimentos de Raper y Thom parecían implicar que, al responder a dos estímulos diferentes, dos especies que viven en el mismo lugar pueden permanecer sin entremezclarse, sin perder, por ende, su identidad. El problema no se agota en dos especies. James C. Cavender, de la Universidad de Ohio, y Raper, que han dedicado mucho tiempo al muestreo de hongos mucosos por todo el mundo, han puesto de relieve que, en ciertos ambientes templados y tropicales, coexisten de cinco a ocho especies, y que no es necesario que todas ellas tengan acrasinas diferentes para mantener su identidad respectiva. Tal como Raper y Thom indicaron en su publicación original, puede haber también "incompatibilidad de superficie celular": algunas especies se agregan juntas en respuesta a la misma acrasina y luego las células de cada una de ellas se separan en el centro de agregación, formando dos individuos pluricelulares distintos. Sin embargo, se han aportado pruebas de la existencia de al menos ocho acrasinas diferentes entre la cincuentena de especies conocidas de hongos mucosos celulares; cuando se hayan investigado todas las especies, con seguridad se pondrá de manifiesto la presencia de nuevas acrasinas.

Ello significa que el cambio evolutivo de los hongos mucosos celulares al pasar de un sistema de quimiotaxis a otro es importante, y plantea una cuestión fascinante: cómo puede alterarse un sistema de señales a partir de un sistema químico de esa naturaleza, con su acrasina, sus receptores para una acrasina específica y, sin duda, con otras sustancias asociadas. Salta a la vista que la presión de selección que actúa sobre los hongos mucosos para que se diversifiquen y ocupen nuevos nichos debe ser intensa. Y, sin negarlo, ¿cómo se realizan los pasos evolutivos necesarios para pasar de un sistema de quimiotaxis con AMP cíclico a un sistema con glorina (o viceversa)?



7. TALLOS Y MASAS DE ESPORAS de *Polysphondylium pallidum* (que como *P. violaceum* segrega y responde a la acrasina glorina), aumentados 16 veces en esta microfotografía obtenida por Raper. Se han desarrollado en el centro de una colonia bacteriana, cuyo borde puede verse en la parte superior de la imagen. Los cuerpos fructíferos aparecen en diferentes estadios del desarrollo: los mayores se están formando todavía y uno que hay a la derecha, con una masa esférica apical de esporas, ya está maduro.

Vale la pena comparar los hongos mucosos del suelo con los sistemas de señales del hombre. Este posee abundante variedad de hormonas y neurotransmisores, que surgieron durante su evolución a partir de animales más sencillos, con muchos menos sistemas de señales. Ello nos conduce a la misma pregunta de antes: ¿cómo surgieron las nuevas hormonas y neurotransmisores a lo largo de los primeros estadios de la evolución animal? Una vez más, podemos suponer que, para mejorar el control, convenía poseer una serie de hormonas y neurotransmisores bien diferenciados, y que, a través de la selección natural, algunos de estos sistemas de señalización dieron lugar a otros nuevos.

Estoy trazando una analogía, burda si se quiere, entre el cuerpo de un mamífero y una parcela de tierra, y postulando para ambos una tendencia hacia la diversificación de las señales químicas. Por ello es concebible que los orígenes de las acrasinas puedan ayudarnos a comprender los orígenes de los sistemas animales, neuronales y hormonales de señalización. Existe, sin embargo, una diferencia de proporción entre los hongos mucosos y los mamíferos: las señales en los animales, seres humanos incluidos, producen un organismo individual dotado de facultades extraordinariamente armonizadas, entre ellas la razón, algo que difícilmente puede esperarse de una palada de suelo.

Control genético de las proteínas del trigo

La calidad tecnológica y nutritiva del grano de trigo depende en gran medida de su composición proteica. Los recientes avances de la biología molecular nos abren nuevas perspectivas en la manipulación genética de dicha composición

Francisco García Olmedo y Pilar Carbonero Zalduegui

La domesticación del trigo, ocurrida hace unos 10.000 años en el suroeste asiático, constituye uno de los hechos seminales de la cultura occidental y su posterior expansión como primera cosecha mundial está íntimamente asociada al progreso del hombre. Esta posición preponderante del trigo se ha debido no sólo a sus buenas propiedades agronómicas –su adaptabilidad y su capacidad productiva en las más variadas situaciones climáticas–, sino también a las peculiares propiedades mecánicas del endospermo, que permiten la elaboración del pan, alimento bíblico que sigue siendo hoy la base de la dieta de una buena parte de la humanidad. El endospermo, que es el tejido mayoritario del grano, constituye un alimento relativamente completo. Contiene almidón y otros hidratos de carbono (60-80 por ciento), proteínas (8-15 por ciento), lípidos (1,5-2,0 por ciento), minerales (1,0-1,5 por ciento) y vitaminas (E, complejo B). La composición proteica del endospermo determina en gran medida la calidad de éste en relación con la elaboración del pan (calidad panadera) e incide de un modo importante en el valor nutritivo de este alimento. En el presente artículo nos proponemos examinar las relaciones entre la composición proteica y la calidad panadera y nutritiva del grano, así como resumir nuestros conocimientos sobre el control genético y la manipulación de dicha composición proteica.

La harina, obtenida por molienda del endospermo, se convierte en pan mediante el proceso de panificación, que comprende tres etapas: amasado, fermentación y cocción. El amasado consiste en el tratamiento mecánico de una mezcla de harina, levadura y agua hasta que la masa adquiere las propiedades viscoelásticas idóneas. La fermentación de la masa de pan es una fermentación alcohólica, cuyo agente es la levadura de panadería (*Saccharomyces cerevisiae*) y cuyos productos principales son anhídrido carbónico y alcohol etílico. El sustrato de la fermentación está constituido por glucosa y maltosa, azúcares fermentables generados por degradación hidrolítica del almidón. La masa debe ser extensible, elástica y suficientemente impermeable a los gases, de tal manera que permita cambios de forma, la retención del CO_2 y el esponjamiento consiguiente durante la fermentación, así como el aumento del volumen del pan en la cocción. La cocción provoca una serie de modificaciones en la masa: (*1*) gelificación del almidón, lo cual aumenta su digestibilidad y la capacidad de retención de agua del pan cocido (hasta 45 por ciento en peso); (*2*) aumento del volumen de la masa esponjosa por expansión del CO_2 y volatilización del etanol (alveolado del pan); (*3*) coagulación de las proteínas, proceso que estabiliza la estructura y el volumen de la pieza de pan; (*4*) muerte de las levaduras e inactivación de los enzimas.

También se obtiene pan de otros cereales: el centeno o la cebada, por ejemplo; de hecho, éste fue el cereal preferido casi hasta el principio de la era cristiana, pero ya Plinio escribió que "el pan de cebada usado en épocas más primitivas había sido condenado por la experiencia...". La capacidad de dar un pan con una alta relación de volumen a peso de harina, con una miga uniformemente alveolada y con unas excelentes propiedades organolépticas singulariza destacadamente al trigo panificable frente al resto de los cereales. Es esta capacidad de la harina de trigo la que depende en gran parte de su composición proteica.

Los estudios sobre las proteínas del endospermo de trigo tienen una larga tradición. Ya en 1745, Beccari describió lo que hoy conocemos como gluten, una masa viscoelástica que se obtiene tras eliminar por lavado el almidón de una masa formada por harina y agua. A lo largo del siglo XIX, Einhof (1805), Taddei (1819) y otros investigadores descubrieron que una fracción importante del gluten, a la que se denomina gliadina, era soluble en mezclas alcohol-agua. A principios del presente siglo, Osborne (1907) propuso una clasificación general de las proteínas basada principalmente en los estudios sobre el endospermo de trigo, que todavía es de uso general.

En la actualidad se consideran los siguientes grupos de proteínas en el endospermo de trigo: albúminas, solubles en agua y en soluciones salinas (0,5 M ClNa); globulinas, solubles en soluciones salinas, pero poco o nada solubles en agua; prolaminas (gliadinas en el caso del trigo), solubles en mezclas alcohol-agua y que requieren la presencia de un reductor de puentes disulfuro para una extracción cuantitativa; glutelinas (gluteninas en trigo), insolubles en los anteriores disolventes, pero solubles en ácidos o bases diluidos; residuo, proteínas no extraídas por ninguno de los disolventes mencionados, usados secuencial o individualmente.

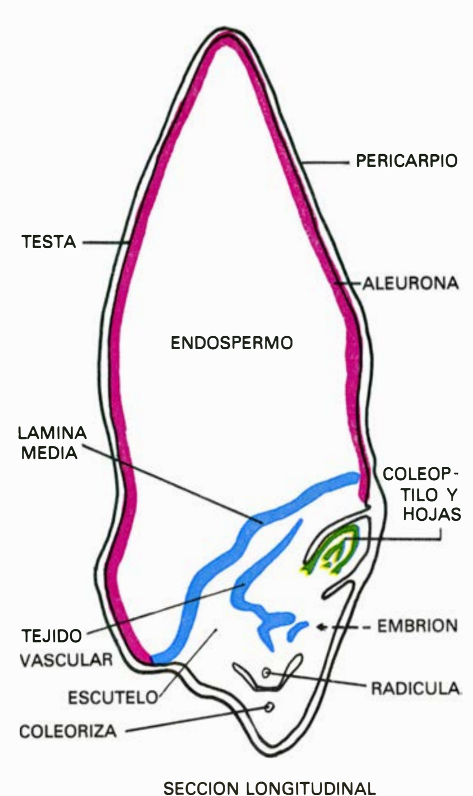
Cada uno de estos grupos está integrado por cierto número de proteínas mayoritarias, muchas de las cuales se han purificado y caracterizado. Además de por sus propiedades de solubilidad, los diferentes grupos pueden distinguirse por otras características: composición en aminoácidos, tamaño y forma de los componentes, localización subcelular, función que desempeñan, etcétera. La fracción mayoritaria está representada por las gliadinas, que son proteínas con una función de reserva, ya que constituyen la principal fuente de nitrógeno para el desarrollo inicial

de la nueva planta durante la germinación del embrión. Las gliadinas se almacenan durante el desarrollo del endospermo en orgánulos especializados denominados cuerpos proteicos o proteosomas. En consonancia con su función de reserva, estas proteínas tienen un elevado contenido en el aminoácido amidado glutamina (por encima del 30 por ciento) y se incrementan preferencialmente cuando se aumenta la proteína del grano, forzando el abonado nitrogenado.

Usando métodos de separación de proteínas de alta resolución, como son los bidimensionales de electroenfoque x electroforesis, se ha puesto de manifiesto que la fracción de gliadinas incluye entre 40 y 50 proteínas. Estas forman familias más o menos relacionadas estructuralmente entre sí y dentro de las cuales existe un alto grado de homología en las secuencias de aminoácidos. Tradicionalmente se distinguen, de mayor a menor movilidad electroforética a pH ácido, las α-, β-, γ- y ω-gliadinas. Más recientemente, algunos autores incluyen en esta fracción prolaminas de alto peso molecular (95.000-136.000 dalton) que son solubles en mezclas de alcohol-agua, una vez desagregadas por reducción de puentes disulfuro [-S-S-Proteína-S-S-Proteína-S-S- + agente reductor X $\rightarrow$ Proteína-S-X]. Estas prolaminas de alto peso molecular (*h*igh *m*olecular *w*eight *p*rolamins = HMWP) se habían considerado antes gluteninas. La distinción entre gliadinas y gluteninas ha dado lugar a frecuentes confusiones debido a que parte de las gliadinas clásicas (α-, β-, γ-, ω-) forman, junto con las HMWP y las gluteninas, agregados de alto peso molecular (10^6 dalton), mediante puentes disulfuro intermoleculares, y no se extraen cuantitativamente por disolventes alcohólicos, si no se rompen antes estos puentes disulfuro.

La calidad del pan depende de la cantidad y calidad de proteína, que, a su vez, depende de la variedad genética y de la nutrición nitrogenada-azufrada de la planta. Con objeto de establecer las relaciones entre composición y calidad tecnológica se han realizado numerosos estudios en los que se han buscado correlaciones entre características del endospermo y características del pan elaborado a partir de él, principalmente la textura de la miga y el cociente entre volumen de pan y peso de harina. Cronológicamente, la primera observación empírica fue la distinción entre trigos o endospermos duros y blandos, y entre vítreos y opacos, lo que permitía hacer ciertas predicciones sobre su comportamiento en la molienda y en la panificación. En general existe una correlación entre dureza y vitrosidad, pero las dos propiedades no parecen tener una base común. Más clara está la relación de vitrosidad con alto contenido en proteína y de opacidad con bajo contenido en proteína. En los trigos blandos, las células se rompen fácilmente en la molienda y se liberan los gránulos de almidón; existe, además, poca adhesión entre los cuerpos proteicos, los gránulos de almidón y los restos de paredes celulares. En cambio, en los trigos duros estos elementos se cementan en una masa más compacta. Aunque caracteres tales como la dureza están notablemente influenciados por las condiciones ambientales y nutricionales, durante el desarrollo y maduración del grano, tienen también una base genética bien establecida.

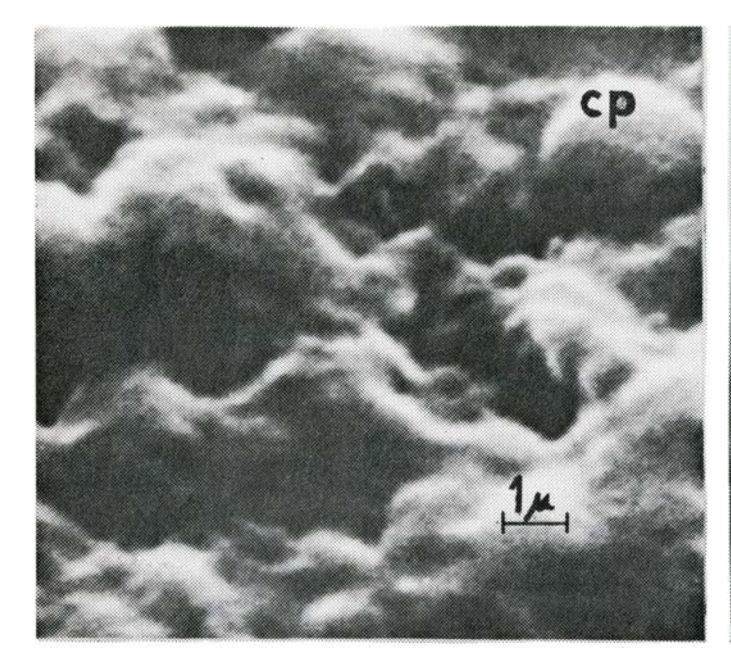
Una segunda correlación fue la observada entre cantidad de gluten y calidad del pan, lo que supuso relacionar por primera vez la composición proteica con la calidad panadera, ya que el gluten se compone principalmente de las fracciones de proteínas que hemos denominado gliadinas, gluteninas y residuo. Esta observación constituyó el punto de partida de estudios encaminados a esclarecer el papel de las proteínas en la panificación y las bases moleculares de las diferencias genéticas de calidad entre una variedad y otra. No se han encontrado correlaciones entre los contenidos de albúminas, globulinas y gliadinas, por un lado, y la calidad tecnológica, por otro. Sin embargo, varios investigadores han hallado correlaciones positivas entre parámetros de calidad y la cantidad de proteína residual que queda después de una extracción exhaustiva con diversos disolventes. Estos resultados sugieren que los agregados de glutenina y residuo, que incluyen las HMWP y parte de las gliadinas unidas por puentes disulfuro, son los principales responsables de las propiedades viscoelásticas de la masa y de la calidad panadera. Un paso importante en la búsqueda de los componentes proteicos individuales con una influencia decisiva en la calidad viene representado por el hallazgo de correlaciones entre diferencias intervarietales en los modelos electroforéticos de las HMWP y ciertos parámetros de calidad. En este caso, la asociación entre proteína individual y parámetro de calidad ha mostrado ser heredable en numerosos cruzamientos. Naturalmente, estas correlaciones no implican relaciones de causa a efecto, pero sí delimitan, junto con los conocimientos que se poseen sobre la bioquímica de las distintas

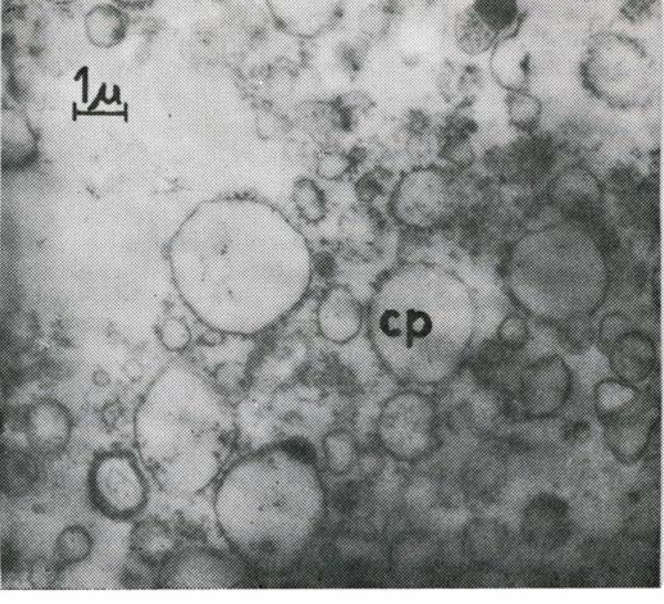


1. GRANO DE TRIGO, en sendos cortes, longitudinal y transversal. De fuera adentro, observamos, en primer lugar, el pericarpio o envoltura del fruto (embrión más endospermo). El endospermo es el tejido mayoritario del grano de los cereales; consta de una capa de aleurona y del endospermo amiláceo rico en gránulos de almidón y proteínas.

proteínas, los posibles modelos que expliquen, por un lado, la estructura de la masa de pan y, por otro lado, las diferencias genéticas de calidad.

Un tipo de prueba más directa viene representada por los ensayos de fraccionamiento y reconstitución, en los que harinas de diferente calidad se fraccionan en sus distintos componentes y a partir de ellos se reconstituyen harinas con componentes intercambiados. Los resultados de estos estudios son coherentes en general con los anteriormente descritos, asignando a las gluteninas, al residuo y, en menor grado, a las gliadinas un papel decisivo en la estructura de la masa. Este planteamiento experimental ha permitido matizar, además, las ideas actuales sobre el papel de





2. CUERPOS PROTEICOS, llamados también proteosomas, son orgánulos aproximadamente esféricos que contienen las proteínas de reserva del grano. Resulta fácil obtenerlos a partir de endospermo en desarrollo, por fraccionamiento subcelular. En las micrografías se muestra una preparación de proteosomas (identificados por *cp*), parcialmente purificados: microscopía electrónica de barrido (*a la izquierda*) y electrónica de transmisión (*a la derecha*).

otros componentes tales como los lípidos y el almidón.

Las características esenciales del modelo más generalmente aceptado sobre la estructura de la masa son las siguientes: (*1*) Las proteínas del residuo, las gluteninas y las HMWP forman una malla tridimensional en la que las distintas subunidades proteicas se unen por puentes disulfuro (-S-S-) para dar largos polímeros lineales con cierto grado de ramificación. Los puentes -S-S-, que originalmente se forman "in vivo", pueden redistribuirse durante el tratamiento mecánico del amasado. Además, pueden formarse "de novo" con posterioridad a la recolección por oxidación de grupos sulfhidrilo (-SH), siendo precisamente el papel probable de ciertos mejorantes de la harina, tales como el bromato o el iodato, el de constituir agentes eficaces de dicha oxidación. (*2*) La red tridimensional se encuentra trabada o cementada por otras proteínas, almidón, lípidos polares y agua, entre las que se dan interacciones hidrofóbicas, iónicas, puentes de hidrógeno, etcétera. Una estructura de estas características tendría las propiedades observadas en la masa: extensibilidad sin rotura, elasticidad para recuperar la forma cuando desaparecen las fuerzas deformantes y capacidad para retener gases y esponjarse.

Las proteínas del trigo son de baja calidad nutritiva. Estudios sobre nutrición humana, realizados a lo largo de la primera mitad de este siglo, han

PROTEINAS	% DE PROTEINA TOTAL	SOLUBILIDAD	AMINOACIDOS (MOLES/100 MOLES)			PESOS MOL. (KILODALTON)	EFECTO CALIDAD		SECUENCIA	OBSERVACIONES
			GLX	PRO	LYS		TECN.	NUT.		
ALBUMINAS	5-8	H_2O y 0,5M ClNa	10-11	7	5	12-24		++		
INHIBIDORES α-AMILASA			11-13	7-8	2,5-5	12,5-60			N-TERMINAL	
TIONINAS			3	4	10	5			COMPLETA	POTENCIALMENTE TOXICA
GLOBULINAS	8-10	0,5M ClNa	11-15	4-5	4-7	10-70		++		
PROTEINAS CM			12-15	10-14	2-3	11-15				SOL. EN CLOROFORMO/ METANOL
GLIADINAS	50-60	ALCOHOL-H_2O	20-50	15-25	0,9		+	−		
LMWG			23-27	9-11	0,3	16-18				SOL. EN CLOROFORMO/ METANOL
α-, β-, γ-			35-40	15-20	0,5	32-44			N-TERMINAL	α ENFERMEDAD CELIACA -S-S- INTRAMOLECULARES
ω-			40-50	20-30	0,3	44-72			N-TERMINAL	NO TIENE -S-S-
GLUTENINAS	10-20	ACIDOS O BASES DIL.	10-15	7-10	2-4		+++	+		-S-S- INTERMOLECULARES
HMWP			30-40	11-14	1,2	95-136	+++		N-TERMINAL	-S-S- INTERMOLECULARES SOL. EN ALCOHOL-H_2O-RSH
RESIDUO	3-6	INSOLUBLE	17	6	6		+++	+		

3. GRUPOS DE PROTEINAS del endospermo de trigo, según la clasificación de Osborne, basada principalmente en criterios de solubilidad. Se consignan también algunos subgrupos de proteínas dentro de estos grandes grupos. En algunos casos, la asignación de los subgrupos no está clara del todo; por ejemplo: las proteínas CM son al menos parcialmente solubles en agua y las prolaminas de alto peso molecular (HMWP, del inglés high molecular weight prolamins) son solubles en alcohol-agua, previa reducción de sus puentes disulfuro -S-S-. Las distintas proteínas difieren en su composición en aminoácidos, presentándose en la tabla sólo los más significativos desde el punto de vista de la clasificación: glutamina + glutámico (GLX) y prolina (PRO), que son los aminoácidos mayoritarios típicos de las proteínas de reserva, y la lisina (LYS), aminoácido esencial limitante. Se han purificado numerosos componentes de los subgrupos y se han determinado sus pesos moleculares y sus secuencias de aminoácidos (totales o parciales). Se consigna también en la tabla la incidencia, en términos semicuantitativos, de los distintos grupos de proteínas sobre la calidad panadera (Tecn.) y nutritiva (Nut.) de la harina.

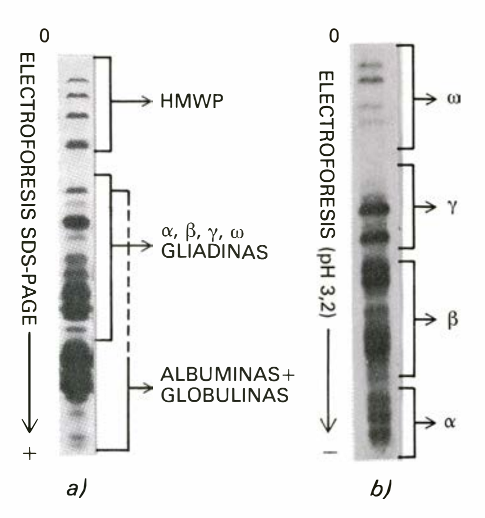
puesto de manifiesto que el requerimiento dietético de ciertos aminoácidos, denominados esenciales, se debe a la incapacidad del organismo humano para sintetizarlos. Varios aminoácidos esenciales están en muy baja proporción en la harina de trigo, siendo la lisina el aminoácido esencial limitante. Por esta razón, el valor biológico de dicha proteína, es decir, el porcentaje de proteína absorbida que es retenida por el organismo, es muy bajo (52 por ciento) si se compara con el de la proteína de huevo (94 por ciento) o el de la proteína de la leche de vaca (90 por ciento). La composición en aminoácidos global de la proteína de trigo es el resultado de la composición de las distintas fracciones y, en último término, de la composición de las proteínas individuales integrantes del conjunto. Concretamente, el bajo contenido global de lisina se debe a que las gliadinas, proteínas de reserva del grano que suponen un 60 por ciento de la proteína total, prácticamente carecen de este aminoacido esencial (por debajo del 1 por ciento), aunque las otras fracciones proteicas, albúminas, globulinas y gluteninas, tienen contenidos más elevados (3-8 por ciento).

Otro aspecto de la calidad nutritiva que debe considerarse es el de la posible presencia en el endospermo de factores antinutritivos. No parece que estos factores sean tan importantes en el caso de los cereales como en el de otros granos, tales como los de las leguminosas, pero se han descrito algunos efectos patológicos asociados a ciertas proteínas de trigo. El mejor estudiado es el conocido como enfermedad celíaca, que consiste en la interferencia con la absorción intestinal de nutrientes producida por componentes de las α-gliadinas en personas sensibles. La sensibilidad es hereditaria, determinada poligénicamente, y da lugar a síntomas adicionales de carácter inmunológico, así como a deficiencias enzimáticas. También se ha descrito recientemente que las tioninas, unas proteínas de pequeño tamaño presentes en el endospermo, tienen efectos citopáticos sobre células de mamíferos en cultivo, incluidas células humanas (HeLa), y bloquean además la síntesis de proteínas y ácidos nucleicos. No tenemos datos de que las tioninas resulten tóxicas por vía oral y es probable que las degraden rápidamente proteasas del tracto intestinal, aunque no se ha descartado todavía que lleguen a afectar a individuos sensibles. Debe mencionarse también que existen en el endospermo inhibidores de enzimas digestivos, tales como proteasas y amilasas, pero no parece que afecten adversamente a la calidad nutritiva de la harina.

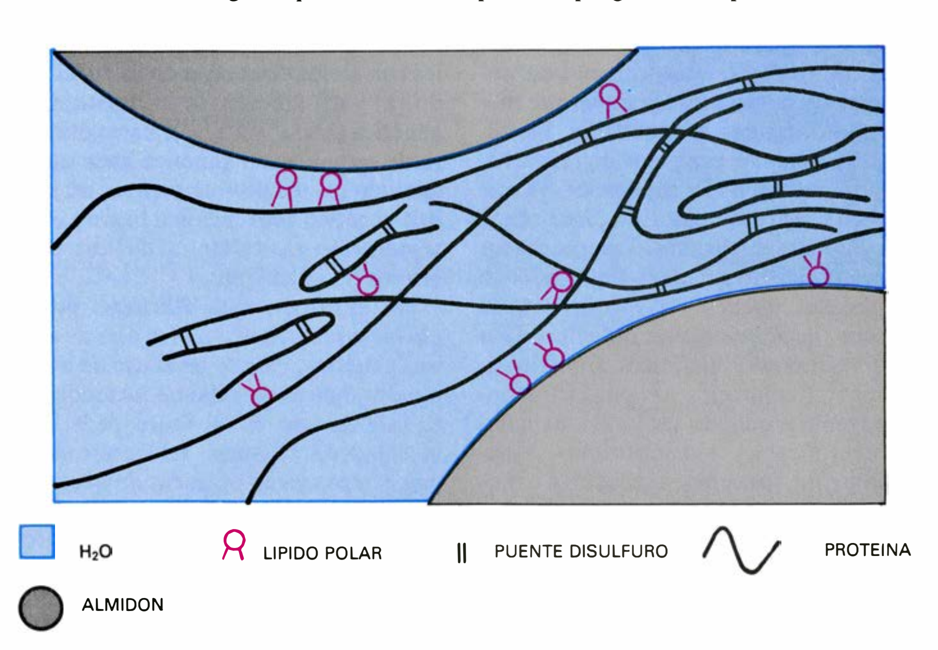
En lo que antecede hemos puesto de manifiesto la estrecha relación existente entre la composición proteica del endospermo de trigo y su calidad tecnológica y nutritiva. De aquí el evidente interés que tiene el estudio del control genético de las proteínas mayoritarias de dicho tejido y su manipulación genética con fines prácticos.

El trigo panificable por excelencia es la especie alohexaploide *Triticum aestivum* L. var. aestivum (genomas AABBDD), que representa la mayor parte del trigo cultivado, siendo mucho menor la producción del alotetraploide *T. turgidum* L. var. durum (genomas AABB), que se usa para elaborar pastas alimenticias [véase "Los recursos genéticos del trigo silvestre", por M. Feldman y E. R. Sears; INVESTIGACIÓN Y CIENCIA marzo, 1981, 50-61]. Los términos alohexaploide y alotetraploide designan los poliploides dotados de seis y cuatro juegos de cromosomas, respectivamente; en la meiosis, no forman multivalentes, sino bivalentes. El término *homeología* (homología ancestral) se utiliza para designar las relaciones de homología entre genomas, cromosomas o genes con un origen común.

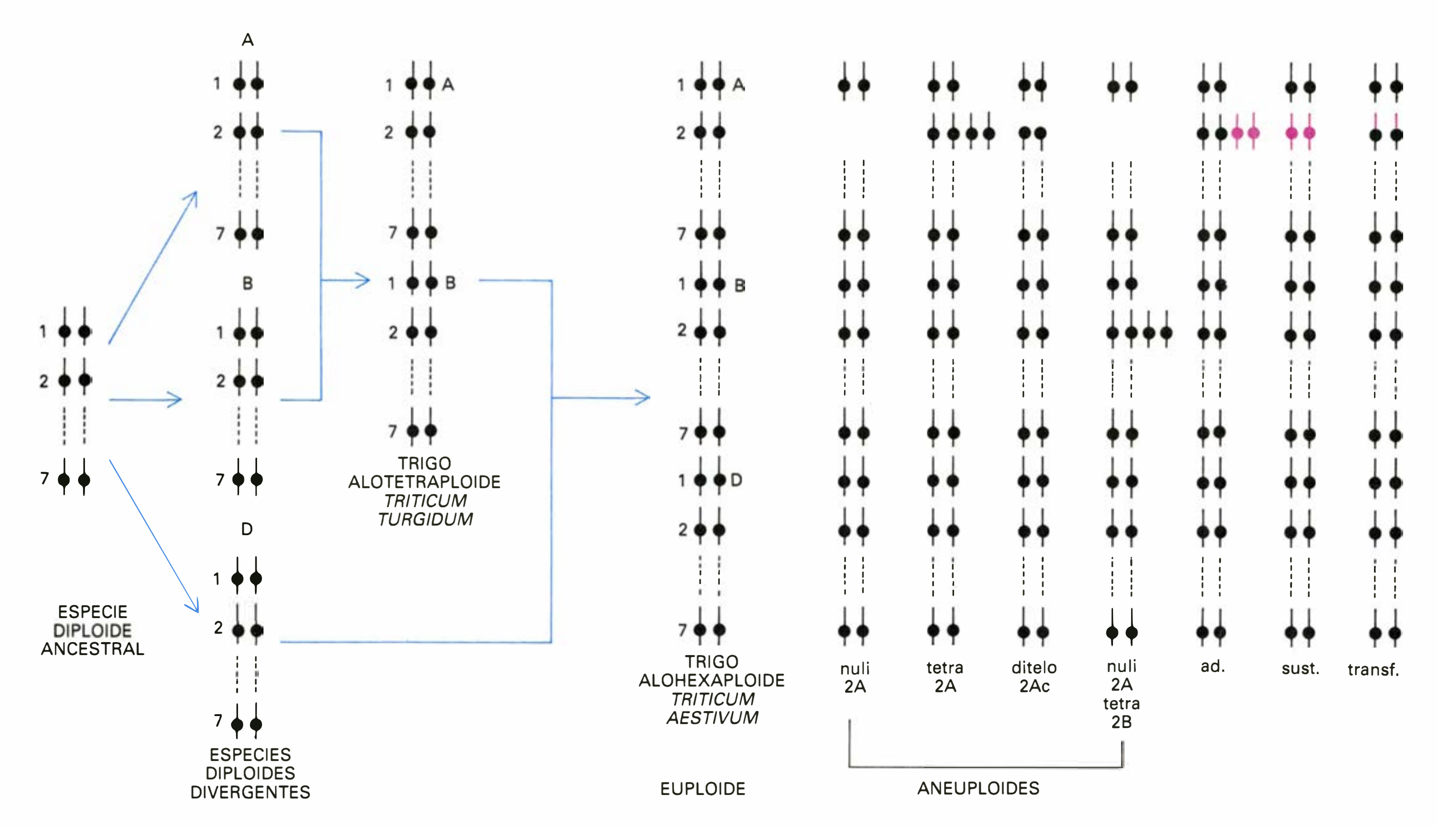
La naturaleza alohexaploide del trigo panificable implica que la mayor parte de la información genética está triplicada en esta especie, existiendo tres *loci* o sitios, localizados en tres pares de cromosomas homeólogos, que codifican para cada carácter cuyo control en una especie diploide correspondería a un solo *locus*. (Y una especie diploide es aquella que tiene dos juegos de cromosomas.) El alohexaploide se ha originado por la integración de tres genomas diploides, donados por tres especies diploides progenitoras procedentes de



4. FRACCIONAMIENTO por electroforesis de las proteínas de endospermo de trigo. A la izquierda, el detergente dodecilsulfato sódico (SDS) extrae la mayor parte de la proteína del endospermo, que puede ser fraccionada por electroforesis en geles de poliacrilamida en presencia del detergente (SDS-PAGE); las proteínas migran tanto más rápidamente cuanto más pequeñas son. Se observa, a la derecha, que el alcohol etílico al 70 por ciento en agua extrae preferentemente las gliadinas, que pueden ser fraccionadas por electroforesis en gel de almidón (*p*H 3,2); las gliadinas migran según su carga positiva a dicho *p*H; se designan α-, β-, γ-, ω-, según su movilidad; cada banda incluye más de una proteína, según se pone de manifiesto por métodos bidimensionales de fraccionamiento.



5. ESTRUCTURA MOLECULAR de la masa de pan. Consiste en una malla tridimensional de proteínas unidas por puentes -S-S-, formada por parte de las α-, β-, γ- y ω-gliadinas, las HMWP y las proteínas del residuo, que incluye los gránulos de almidón (más o menos intactos durante la fermentación y gelificados en la cocción), los lípidos polares y otras proteínas como elementos de relleno y trabado. Los lípidos polares desempeñan un papel importante en las propiedades viscoelásticas de la masa: sirven de puente entre dominios hidrofílicos del almidón y de algunas proteínas y dominios hidrofóbicos de las proteínas que forman la malla. (Lípidos polares son los dotados de un extremo hidrofílico y otro hidrofóbico.)



6. CONSTITUCION CROMOSOMICA de los trigos cultivados. El trigo tetraploide *Triticum turgidum* L. se originó de dos especies diploides, con siete pares de cromosomas cada una, procedentes a su vez de una especie ancestral común; una forma silvestre del trigo diploide *T. monococcum* L. aportó el genoma A y un *Aegilops* diploide no identificado aportó el B. Se usa *T. turgidum* var. durum para elaborar pastas y *T. monococcum* ya no se cultiva. El trigo hexaploide, *T. aestivum* L., usado en panadería, se originó del trigo tetraploide y de la especie diploide *Aegilops squarrosa* (syn. *T. tauschii*), que aportó el genoma D. A partir del cultivar hexaploide "Chinese Spring" se han obtenido líneas en las que la constitución cromosómica normal (euploide) está alterada con respecto a un cromosoma o brazo cromosómico (aneuploides) por adición o sustitución de cromosomas de especies o géneros distintos.

una especie ancestral común. Así, los 21 pares de cromosomas del trigo panificable forman tres genomas homeólogos (AA, BB, DD), cada uno con 7 pares de cromosomas (1A-7A, 1B-7B, 1D-7D). Se dan varias situaciones posibles tanto para los genes estructurales que codifican para las proteínas del endospermo como para los genes que modifican o regulan su expresión. No todos los sistemas genéticos de un alohexaploide poseen *loci* triplicados, ya que se han observado casos de *locus* sencillo y de *loci* duplicados. Los productos correspondientes a *loci* duplicados o triplicados pueden ser idénticos o, al menos, operativamente indistinguibles por los métodos analíticos disponibles, o, por el contrario, los productos homeogénicos pueden ser individualmente identificables y cuantificables. Análogamente, los genes reguladores o modificadores duplicados o triplicados pueden tener una acción equivalente, o, por el contrario, cada uno de los homeogenes reguladores ha podido diversificarse, ejerciendo una acción diferente o incluso afectando a genes estructurales distintos.

Además de la complejidad genética que se deriva de la naturaleza alohexaploide del trigo, hay que señalar que el caso del endospermo se complica aún más debido a la naturaleza triploide de este tejido (AAA, BBB, DDD). De las tres copias de cada cromosoma distinto presentes en una célula de endospermo, dos son de aportación materna y uno de procedencia paterna, por lo que cruzamientos recíprocos entre parentales con alelos diferentes en un *locus* dado (a^1 y a^2) difieren en su constitución genética ($a^1a^1a^2$ y $a^2a^2a^1$). Esta considerable redundancia genética hace que a menudo no sea posible realizar un análisis genético convencional basado en la segregación de variantes alélicas o de grupos de ligamiento.

Sin embargo, esta dificultad puede obviarse recurriendo a otro tipo de análisis genético, basado en el uso de aneuploides, que es especialmente asequible en esta especie. E. R. Sears, de la Universidad de Missouri, ha desarrollado una extensa gama de líneas de trigo que se denominan aneuploides, en las que faltan cromosomas, brazos o segmentos cromosómicos, así como líneas en las que la dotación cromosómica normal (euploide) del trigo se ha aumentado específicamente para un cromosoma dado o sustituido por cromosomas, brazos o segmentos cromosómicos de especies distintas. El disponer de series completas de los distintos tipos de líneas permite explorar sistemáticamente todos los genomas, anulando (o reforzando) bloques de información genética (cromosomas, brazos o segmentos cromosómicos) y correlacionando dichas alteraciones genéticas con la falta de expresión (o el reforzamiento) de caracteres observables.

Describiremos brevemente los distintos tipos de líneas usados en este modo de análisis genético. Se llaman líneas monosómicas y nulisómicas las que carecen respectivamente de uno o de los dos cromosomas de uno de los 21 pares de cromosomas que constituyen el euploide; monotelosómicas o ditelosómicas, si carecen respectivamente de uno o de dos brazos cromosómicos homólogos; trisómicas o tetrasómicas, cuando poseen, respectivamente, una o dos dosis extra de un cromosoma dado. Las líneas monosómicas, monotelosómicas y trisómicas segregan y las tetrasómicas son inestables cuando se autofecundan, por lo que es preciso realizar recuentos cromosómicos para identificar en la descendencia qué plantas mantienen las características descritas. La mayoría de las plantas nulisómicas tienen una fertilidad reducida y escaso vigor. De fertilidad y vigor normales son las líneas denominadas nuli-tetrasómicas compensadas, que

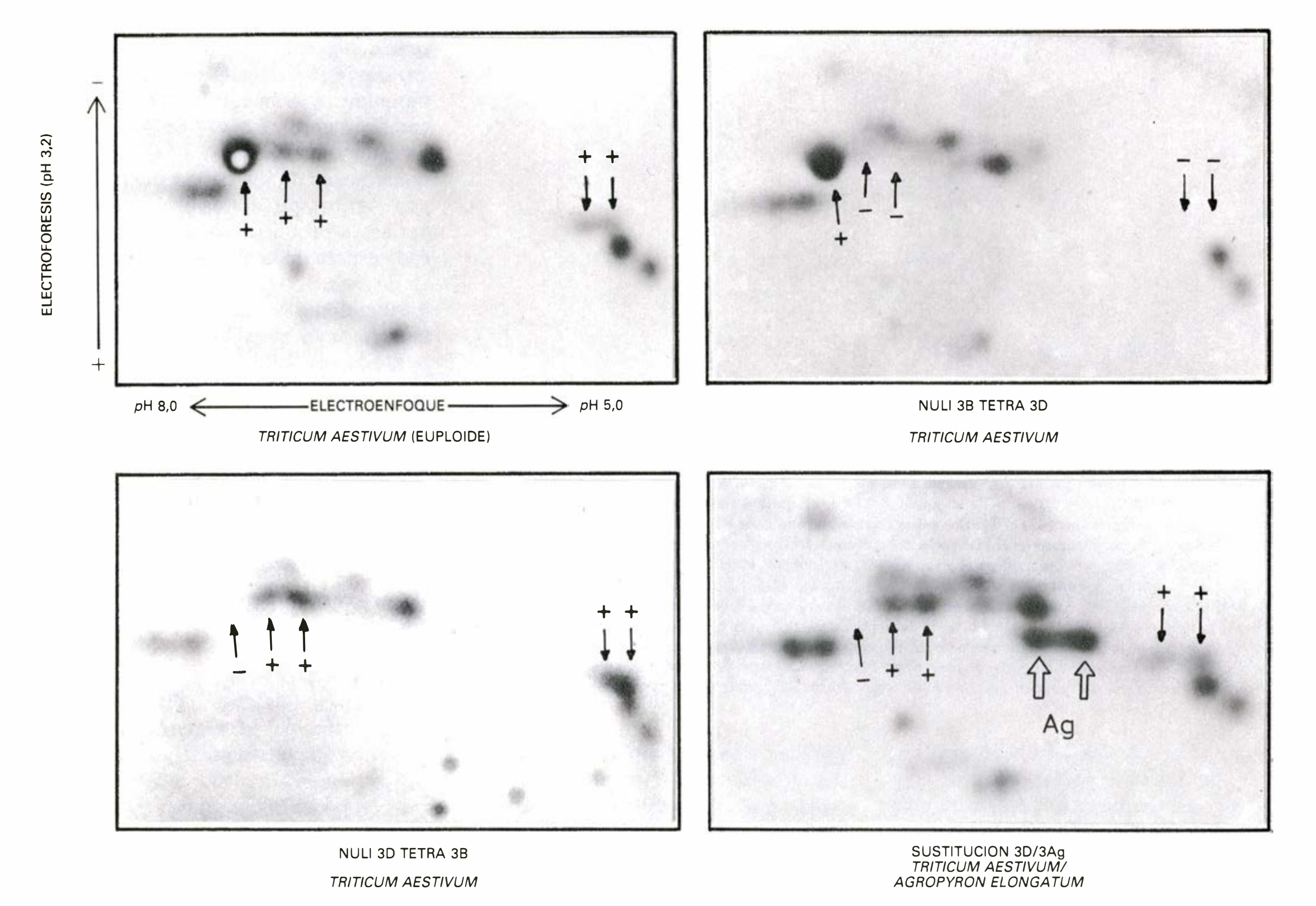
son nulisómicas para un cromosoma y tetrasómicas para uno de sus homeólogos (por ejemplo, nuli 2A-tetra 2B, nuli 2A-tetra 2D; nuli 4D-tetra 4A, nuli 4D-tetra 4B).

Se han obtenido además otros tipos de stocks genéticos de gran utilidad en el presente contexto: las líneas de adición, sustitución y transferencia. En las primeras se han añadido uno o dos cromosomas ajenos; en las segundas se ha sustituido un par de cromosomas por un par homeólogo, de otra especie o género, o por un par homólogo de otra variedad de la misma especie y, finalmente, las líneas de transferencia llevan un segmento cromosómico homeólogo de una especie distinta como sustituyente en un par de cromosomas.

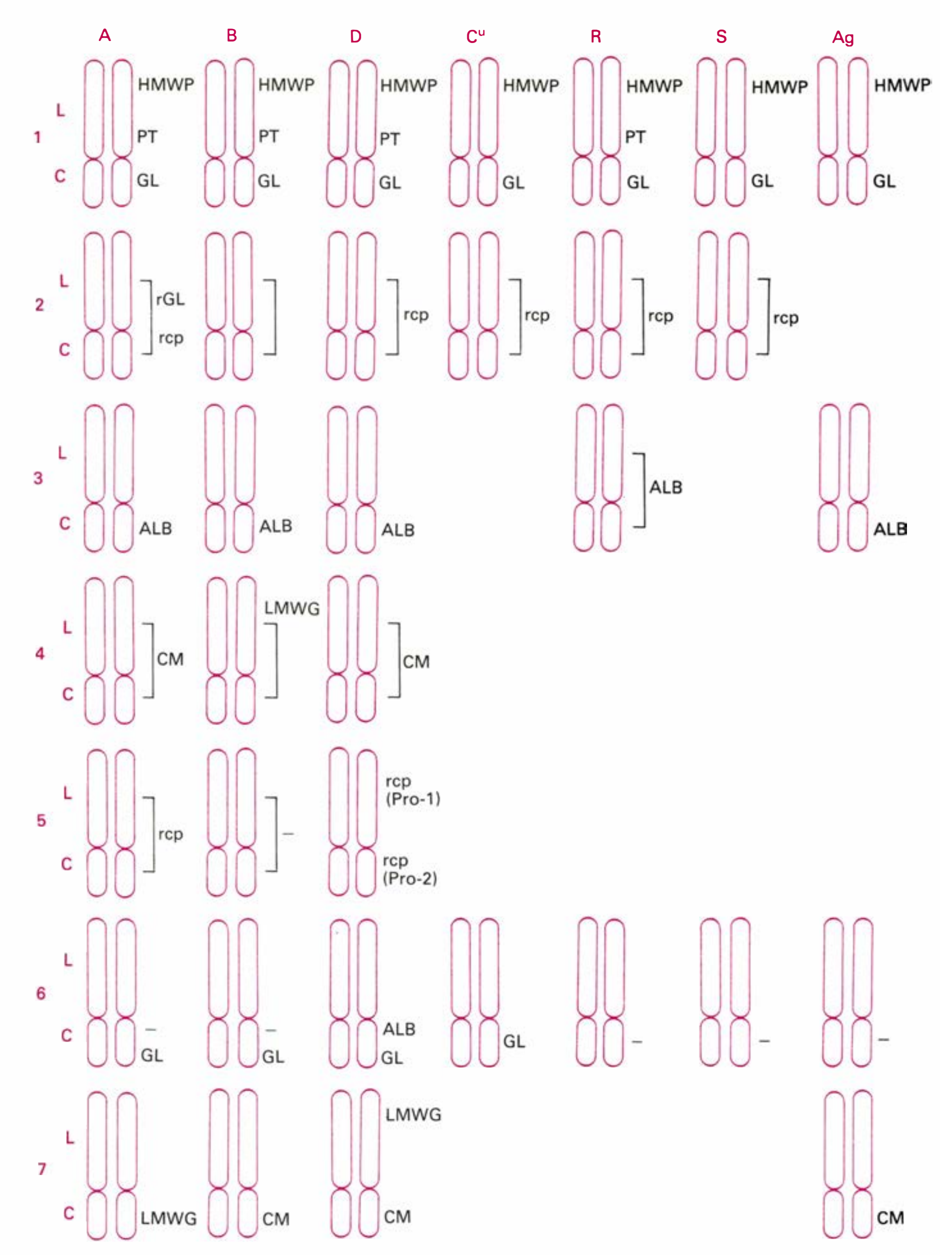
Se aplican tres criterios para asignar los genes que codifican para las distintas proteínas del endospermo a segmentos cromosómicos concretos: (*1*) concomitancia de la desaparición de la proteína con la ausencia de un cromosoma o un segmento cromosómico, (*2*) variación cuantitativa del nivel de expresión en función de la variación de la dosis cromosómica o del número de copias de un segmento que incluya el gen estructural; (*3*) aparición de la proteína en líneas en las que se ha añadido o sustituido un cromosoma, o se ha transferido un segmento cromosómico. El análisis de series completas de líneas nulisómicas, nulitetrasómicas y ditelosómicas aporta las pruebas del primer tipo. Los efectos de dosis génica pueden ser investigados en las series monosómicas, trisómicas, tetrasómicas y nuli-tetrasómicas. Esta última serie es particularmente útil, no sólo porque aporta datos de los dos primeros tipos, sino porque sus fenotipos son generalmente vigorosos y fértiles, siendo fácil su mantenimiento y propagación. El tercer tipo de pruebas se obtiene analizando sustituciones homólogas intraespecíficas, sustituciones homeólogas interespecíficas e intergenéricas, adiciones monosómicas y disómicas y líneas de transferencia. Este tipo de corroboración resulta particularmente útil en estudios con proteínas, porque permite distinguir entre genes estructurales y genes modificadores o reguladores.

El método de análisis genético que acabamos de describir ha permitido asignar a brazos cromosómicos concretos la mayoría de los genes estructurales correspondientes a proteínas abundantes y específicas del endospermo, así como localizar los genes homeólogos correspondientes en especies silvestres y cultivadas más o menos distantes del trigo, como son *Aegilops umbellulata* (= *Triticum umbellulata*), centeno (*Secale cereale*), cebada (*Hordeum vulgare*) y *Agropyron elongatum*.

Los genes estructurales que codifican para las α-, β-, γ- y ω-gliadinas, principales proteínas de reserva del grano, se localizan en los brazos cortos de los cromosomas de los grupos homeólogos 1 y 6 de los genomas (ABD) del trigo. Tanto los datos de la bioquímica (composición y secuencia de aminoácidos) como los aportados por la segregación en cruzamientos apropiados indican que dichos genes forman al me-



7. USO DE ANEUPLOIDES para localizar los genes que determinan distintas proteínas. Las proteínas se extraen del endospermo y se fraccionan por un método bidimensional de electroenfoque (primera dimensión) × electroforesis (segunda dimensión). El electroenfoque separa las proteínas según su punto isoeléctrico y la electroforesis según su carga al *p*H elegido; cada mancha del mapa bidimensional representa una sola proteína cuya presencia (+) o ausencia (−) en distintas líneas aneuploides indica la localización cromosómica del gen estructural que la codifica. En el ejemplo elegido se determina la localización de los genes que codifican para ciertas albúminas en cromosomas del grupo 3 y se demuestra la expresión de genes que codifican para albúminas homeólogas de *Agropyron elongatum* (Ag) en una línea de sustitución 3D/3Ag. (Estas fotografías se han tomado de la tesis de M. Rodríguez-Loperena.)



8. LOCALIZACION CROMOSOMICA DE LOS GENES que codifican, modifican o regulan las proteínas mayoritarias del endospermo de trigo y de especies afines. Para una mejor comprensión, los símbolos empleados en el mapa genético de arriba no son los correspondientes a los genes, sino a las proteínas pertinentes. En casi todos los casos, el símbolo representa varias proteínas estrechamente relacionadas y, por tanto, varios genes: prolaminas de alto peso molecular (HMWP); tioninas (también denominadas purotioninas en el trigo, PT); α-, β-, γ- ω-gliadinas (GL); albúminas (ALB); proteínas CM (CM); gliadinas de bajo peso molecular (LMWG). Los genes con efecto regulador o modificador del contenido proteico total (rcp) y de las gliadinas (rGL) también se ilustran de forma no convencional. No se han representado otros muchos genes correspondientes a caracteres bioquímicos, morfológicos o fisiológicos cuya localización cromosómica se conoce en la actualidad. Los genomas A, B y D pertenecen al trigo hexaploide *T. aestivum* y la numeración cromosómica empleada (1-7; brazo largo, L, o brazo corto, C) es la correspondiente a esta especie. En las especies *Aegilops umbellulata* (C^u), centeno (R), cebada (S) y *Agropyron elongatum* (Ag) se han representado sólo los cromosomas involucrados en el control genético de las proteínas o efectos investigados en cada una de ellas; las numeraciones de los cromosomas son las que les corresponderían por su homeología con los de trigo, no las asignadas originalmente para cada especie.

nos 11 agrupaciones en tándem. El hecho de que una especie próxima, como *Aegilops umbellulata*, tenga también los genes que codifican para estas gliadinas en cromosomas correspondientes a los grupos de homeología 1 y 6, mientras que especies más distantes, como centeno, cebada o agropyron, los tengan sólo en el cromosoma correspondiente al grupo 1, sugiere que los genes localizados en el grupo 6 se originaron por una translocación de un segmento del cromosoma 1 al cromosoma 6, ocurrida en un progenitor de las especies *Aegilops-Triticum*, con posterioridad a las ramificaciones que conducen a *Secale, Hordeum* y *Agropyron*. (Se llama translocación al cambio de posición de un segmento de un cromosoma a otra parte del mismo o de diferente cromosoma.) Además, todos los cromosomas del grupo 1 examinados tienen genes que determinan las proteínas denominadas HMWP, localizados en sus brazos largos. Ya hemos indicado que esta fracción, que comprende 3-5 componentes y forma parte de los cuerpos proteicos, presenta cierta variabilidad genética intraespecífica que parece estar asociada a las diferencias intervarietales en calidad panadera. La similitud estructural de los cromosomas del grupo 1 en todas las especies estudiadas se pone de manifiesto por la distribución de genes que determinan las tioninas y que están localizados en los brazos largos de los cromosomas de este grupo.

En algunos casos se ha podido alcanzar mayor detalle en el mapa genético mediante el análisis de colecciones de líneas de transferencia con cromosomas recombinantes *Agropyron*-trigo, en los que un segmento terminal de un cromosoma de *Agropyron* reemplaza a segmentos homeólogos de un cromosoma de trigo. Por métodos citogenéticos (frecuencia de apareamiento meiótico del cromosoma recombinante con el cromosoma del trigo intacto) es posible ordenar las líneas de transferencia según el tamaño del segmento cromosómico transferido; por análisis bioquímico, se averigua si los genes que codifican para un cierto tipo de proteína están o no localizados en segmentos homeólogos de los cromosomas de trigo y *Agropyron* (mutua exclusión de las variantes proteicas correspondientes a cada especie indica homeología); a partir de ambos, obtendremos el orden de localización de distintos genes en el brazo cromosómico investigado. De este modo se ha demostrado que los genes que codifican para albúminas, localizados en los brazos cortos de los cromosomas del grupo 3, ocupan posiciones homeólogas, próximas al centrómero, en los cromosomas 3D y 3Ag. Del mismo modo se han podido asignar posiciones homeólogas en los brazos cortos de los cromosomas 7D y 7Ag para los genes que rigen determinadas globulinas hidrófobas, denominadas proteínas CM.

En general, tanto la localización de genes que codifican para proteínas mayoritarias del endospermo como la de genes que determinan otras proteínas no aludidas en este artículo (enzimas, por ejemplo) indican una considerable conservación de los mapas génicos en los distintos genomas, a pesar de su evolución divergente, aunque también ponen de manifiesto algunos cambios estructurales en los cromosomas. Por ejemplo, la localización de genes para proteínas CM (cromosoma 4A; brazos cortos de 7B y 7D) y para

gliadinas de bajo peso molecular LMWG (brazo corto de 7A; 4B; brazo largo de 7D) sugiere la existencia de translocaciones recíprocas entre cromosomas de los grupos 4 y 7, que han ocurrido en las especies parentales de *T. aestivum* o durante la evolución posterior al origen de dicha especie.

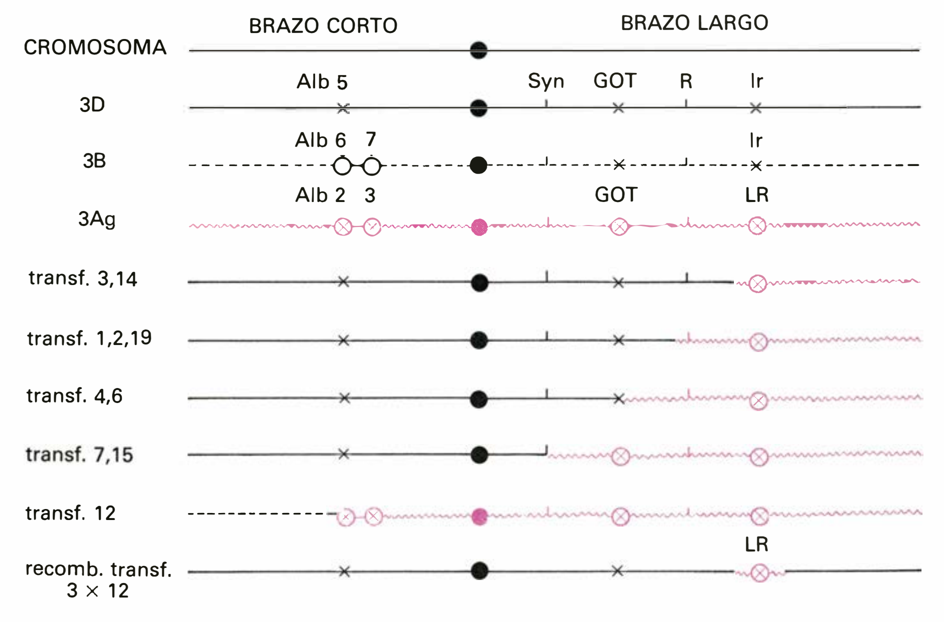
El aumento del contenido en proteínas del grano y de la proporción de aminoácidos esenciales –especialmente lisina– en la proteína, manejando los factores genéticos y ambientales, son objetivos agronómicos de indudable interés, pero de difícil consecución. Existe una correlación negativa entre rendimiento en grano por unidad de superficie y contenido porcentual en proteínas del mismo, aunque un aumento en el rendimiento de la cosecha suele conducir a un aumento de la producción de proteína por unidad de superficie. Existe, además, otra correlación negativa entre proteína del grano y proporción de lisina en la proteína, pero de nuevo un incremento en el contenido proteico del grano lleva consigo un incremento en el contenido en lisina referido al peso de grano, si bien este incremento es proporcionalmente menor.

Son numerosos los investigadores que no han tenido en cuenta estas correlaciones, por lo que la literatura científica relativa al control genético de la cantidad y calidad de proteína del grano es confusa. Sólo disponemos de pruebas sólidas para los efectos genéticos asociados a cromosomas de los grupos 2 y 5. Se ha observado que cuando un par de cromosomas 2R de centeno sustituye a cualquiera de los pares de cromosomas 2 de trigo (2A, 2B, 2D), el contenido en proteínas del grano aumenta considerablemente. Esta observación dio lugar a un estudio posterior realizado en el Plant Breeding Institute (Cambridge, Inglaterra) en el que se investigaron los efectos, sobre el contenido en proteínas de los cromosomas del grupo 2 de las especies (*Aegilops umbellulata* (C^u), *Ae, comosa* (M) y *S. montanum* (R^m) en comparación con cromosomas del grupo 2 de trigo (A, B, D). Analizando líneas de sustitución, se ordenaron los cromosomas según su capacidad para promover la producción de proteína en el grano: $2A > 2M > 2B > 2C^u > 2D > 2R^m$. En confirmación de estos resultados, la sustitución 2M/2D en dos cultivares de trigo comerciales dio lugar a incrementos en el contenido en proteínas del grano.

El cultivar Atlas 66 tiene un contenido en proteínas excepcionalmente elevado y por esta razón se ha puesto gran empeño en analizar la base genética de dicho carácter y en transferirlo a cultivares con buenas propiedades agronómicas. Analizando líneas de sustitución intraespecíficas de Atlas 66 sobre el cultivar Chinese Spring, Rosalind Morris y sus asociados, de la Universidad de Nebraska, localizaron los genes responsables del alto contenido en proteína principalmente en el cromosoma 5D; un efecto menor ejercían genes localizados en el cromosoma 5A. Más tarde se identificaron, por análisis genético convencional, dos genes, *Pro1* y *Pro2*, localizados respectivamente en los brazos largo y corto del cromosoma 5D, responsables del efecto positivo sobre el contenido en proteína.

Los efectos descritos son generales; afectan, al parecer, a la totalidad de la proteína del grano. La alteración de la composición proteica de éste implica alterar las proporciones relativas de las distintas proteínas o de grupos de proteínas concretos. Por ejemplo, un bloqueo parcial de la síntesis de gliadinas, que prácticamente carecen de lisina, y un incremento de las otras fracciones proteicas daría lugar a un incremento en el contenido del grano en el aminoácido esencial limitante. Poco ha aportado todavía el análisis de aneuploides a la localización de efectos modificadores o reguladores específicos. En todas las proteínas investigadas se ha observado una variación lineal de la cantidad de proteína en función de la dosis génica. Para algunas proteínas codificadas por sistemas genéticos no triplicados (duplicados o sencillos) se ha demostrado que la respuesta a la dosis génica (cantidad de proteína por número de genes) es controlada por genes localizados en cromosomas distintos de aquellos donde se localizan los genes estructurales para dichas proteínas. También se ha encontrado que genes localizados en el cromosoma 2A pueden bloquear la expresión de dos genes localizados en el cromosoma 6D, que determinan gliadinas, en líneas tetrasómicas para el cromosoma 2A. El conocimiento de estos efectos específicos es todavía demasiado rudimentario para que pueda dar lugar a resultados prácticos.

Puede estimarse que las proteínas correspondientes a los genes estructurales que ya han sido asignados a sus correspondientes cromosomas representan más del 80 por ciento del contenido proteico del endospermo. Se sabe que muchos de estos genes forman grupos de repeticiones en tándem, por lo que no debe considerarse inviable la alteración de la composición proteica del endospermo mediante inserción o delecciones de segmentos cromosómicos, utilizando las técnicas que algunos han denominado como "ingeniería cromosómica". Aunque el trigo panificable es un hexaploide, se comporta como un diploide, es decir, los cromosomas homeólogos ni se aparean en la meiosis ni se recombinan normalmente, a pesar de su similitud estructural derivada de



9. TRANSFERENCIA GENETICA EXTRAESPECIFICA desde la especie silvestre *Agropyron elongatum* hasta el trigo. El procedimiento seguido para introducir un segmento interno del cromosoma 3Ag (*A. elongatum*), portador de un gen determinado, en el cromosoma homeólogo 3D (*T. aestivum*) comprende las siguientes etapas: en primer lugar, obtención de la línea de sustitución 3D/3Ag; en segundo lugar, formación de recombinantes 3Ag-3D por cruzamiento de la línea de sustitución con el trigo en condiciones en las que se bloquea la acción del gen diploidizante (Ph), lo que promueve el apareamiento y recombinación entre cromosomas homeólogos; en tercer lugar, caracterización de los recombinantes mediante el estudio del apareamiento meiótico con cromosomas 3Ag y 3D intactos y el análisis de marcadores bioquímicos de dichos cromosomas; finalmente, nuevo ciclo de recombinación entre cromosomas recombinantes apropiados (en el ejemplo representado aquí: recombinación entre las línes de transferencia 3 y 12).

su origen ancestral común. Esta falta de apareamiento y recombinación homeóloga está bajo el control positivo de un gen diploidizante (*Ph*), localizado en el brazo largo del cromosoma 5B. Existen distintos procedimientos de manipular el sistema genético *Ph* para obtener cromosomas recombinantes donde se hayan sustituido distintos segmentos por segmentos homeólogos procedentes de especies cultivadas o silvestres más o menos distantes del trigo. Por recombinación entre líneas de transferencia apropiadas surgen nuevos cromosomas recombinantes en los que el segmento homeólogo transferido quede reducido a un mínimo que incluya los genes que deseábamos introducir en el trigo.

La estrategia que acabamos de describir ha dado resultados prácticos satisfactorios en la transferencia de genes que confieren resistencia a enfermedades desde especies silvestres al trigo y creemos que puede dar resultados igualmente positivos en la modificación genética de la composición proteica del endospermo.

La utilización de las modernas técnicas de recombinación de ADN "in vitro" –lo que se ha denominado ingeniería genética– en la mejora de las plantas cultivadas supondrá la ruptura de las barreras que limitan el flujo de material genético en los tipos de manipulaciones que acabamos de describir. En principio, la ingeniería genética posibilitará la transferencia de genes desde otros organismos hasta las plantas cultivadas. El endospermo de los cereales es el principal producto agrícola: representan casi las tres cuartas partes de la superficie sembrada. La nueva tecnología debe permitir tanto "perfeccionar" este tejido para los fines tradicionales a que se ha venido destinando, como hacer expresarse en él nueva información genética que posibilite utilizaciones hasta ahora insospechadas. En el caso concreto del trigo, se está abordando la clonación de los genes que codifican para las proteínas mayoritarias del endospermo. Se trata de una etapa obligada para la caracterización y manipulación "in vitro" de estos genes en relación con los objetivos antes mencionados. Otra etapa importante a remontar será la concerniente a la obtención de vectores apropiados para reintroducir en la planta los genes alterados "in vitro", problema éste todavía sin resolver para los cereales, pero el hecho de que se hayan encontrado soluciones incipientes para otras especies vegetales hace concebir esperanzas de encontrar soluciones a medio plazo.

Temas metamágicos

Donde un discurso sobre el Lisp-uaje concluye con un festín pantagruélico a la usanza italiana

Douglas R. Hofstadter

Di el mes pasado una descripción de la torre de Brahma, rompecabezas ideado por Édouard Lucas, cuyo objetivo es transferir una torre de 64 discos de oro, ensartados en una aguja, hasta otra aguja, auxiliándose para ello de una tercera aguja en la cual podemos situar discos provisionalmente. Es preciso ir tomando los discos de uno en uno, con la condición adicional y única de que nunca puede quedar un disco descansando sobre otro de diámetro menor. El problema que dejé al cuidado de los lectores consistía en dar una descripción recursiva, expresada en el lenguaje de ordenador llamado Lisp, de cómo conseguir tal objetivo (y consiguientemente, de consuno con los sacerdotes de Brahma, dar fin al mundo).

Ya señalé entonces que la recurrencia era bastante evidente: para saber cómo transferir 64 discos de una aguja a otra es suficiente saber cómo hacer lo mismo con 63. Para recapitular la idea, héla aquí: supongamos que la torre de Brahma, con sus 64 discos, se encuentra inicialmente apilada en torno a la varilla *a*. En la parte superior de la figura 1 vemos un dibujo esquemático donde se ha representado la totalidad de los 64 discos por tan sólo cuatro. Antes de nada, basándonos en nuestra hipotética capacidad para trasladar 63 discos, llevemos 63 discos desde la aguja *a* hasta la aguja *c*, valiéndonos de la aguja *b* como "auxiliar". El aspecto que presenta esta situación puede verse en segunda posición desde lo alto de la figura. (Obsérvese que 4 hace el papel del 64 de mi dibujo primitivo, así que el 3 hace el papel del 63; en cambio, por alguna razón, el 1 no hace el papel del 61. ¿No le resulta curioso?) Ahora, sencillamente, alcemos el disco todavía restante en *a* –el mayor de todos– y enfilémoslo en la aguja *b*, como vemos en la tercera ilustración desde arriba.

Ya nos damos cuenta ahora de cuán fácil será terminar: bastará, sencillamente, sacar partido de nuestra capacidad para transferir 63 discos desde *c* hasta *b*, usando esta vez la aguja *a* como auxiliar. Debe observarse que en esta ocasión *a* tiene el papel auxiliar que antes le fuera encomendado a *c* durante la maniobra precedente de transferencia de los 63 discos. Una fracción de segundo antes de que el último disco sea colocado en su lugar, la situación es la que vemos en la parte más baja de la ilustración. ¿Por qué antes de quedar en su sitio? ¿Por qué no después? Porque el mundo entero se encontraría para entonces reducido a polvo, ¡y es muy difícil dibujar polvo!

Claro, alguien podría quejarse de que yo he dejado de lado las partes más difíciles. "Usted supone que por arte de magia nosotros tenemos capacidad para transferir 63 discos." Aunque así pueda parecer, tal suposición nada tiene de mágico. Después de todo, para poder trasladar 63 discos bastará saber cómo trasladar 62. Y para trasladar 62 es suficiente saber trasladar 61. Se va así descendiendo, peldaño tras peldaño, hasta salir por la planta baja, esto es, por el "caso embrionario" del problema de la torre de Brahma, que corresponde al caso de un solo disco. Admito, desde luego, que es preciso ir llevando la cuenta del punto del proceso donde nos encontramos, lo cual puede ser un poco tedioso, pero es pura cuestión de "contabilidad" por así decirlo. Al menos en teoría se podría ahora llevar a efecto el proceso completo... ¡si es que se sienten deseos de ver el fin del mundo!

Como primera aproximación al problema de redactar en Lisp la función requerida, demos una descripción verbal del proceso. Llamemos a las tres varillas "vo", "vd" y "va", abreviaturas de "varilla origen", "varilla destino" y "varilla auxiliar". Vamos allá:

Para trasladar una torre de altura *n*
desde vo hasta vd, con ayuda de va:
si n = 1,
basta llevar directamente ese
único disco desde vo hasta vd;
en cualquier otro caso,
ejecútense los tres pasos siguientes:
(1) traslade una torre de altura
n − 1 desde vo hasta va
con auxilio de vd;
(2) lleve 1 disco desde vo hasta
vd;
(3) traslade una torre de altura
n − 1 desde va hasta vd,
con ayuda de vo.

Los dos pasos aquí numerados 1 y 3 son las dos etapas de recurrencia; parecen contornear la paradoja y depender de la capacidad que están ayudando a definir. El rasgo salvador es que en lugar de aludir a n aluden tan sólo a n − 1. Obsérvese que en el paso 1, va hace el papel de "punto de destino", mientras que vd tiene el de "auxiliar"; en el paso 3, va desempeña la función de "origen" y vo el de "auxiliar". Dado que el proceso completo es recursivo, cada aguja irá cambiando de sombrero muchas veces a lo largo del proceso de transferencia. Esta es precisamente la belleza del rompecabezas, y, en cierto modo, también es esta la belleza de la recurrencia.

¿Cómo efectuar la transición de español al Lisp? Es muy sencillo:

```
(def traslatorre
 (lambda (n vo vd va)
 (cond
  ((eq n 1) (traslaundisco vo vd))
  (t (traslatorre (sub 1 n) vo va vd)
     (traslaundisco vo vd)
     (traslatorre (sub 1 n) va vd vo)))))
```

¿Dónde se encuentran en Lisp los equivalentes de "desde", "hasta" y "con auxilio de"? Parecen haber desaparecido. ¿Cómo podrá entonces el Lisp-genio saber a qué aguja ha de ir en cada momento encomendando cada papel? La respuesta es que tal información le es aportada *posicionalmente*. Hay en esta definición funcional cuatro parámetros: un entero y tres "varillas ficticias". La primera de las tres varillas hace de varilla origen, la segunda, de destino, y la tercera, de auxiliar. Por ello, en la lista inicial de parámetros, que va inmediatamente tras "lambda", se encuentran en orden "vo vd va". En la primera llamada recursiva, esto es, en la traducción a Lisp del paso 1, se encuentran en el orden "vo va vd", lo que indica que va y vd han cambiado entre sí de sombrero. En la segunda llamada recursiva, que es la traducción a Lisp del paso 3, podemos ver que ahora va y vo han intercambiado sus papeles.

La idea fundamental es que los nombres atómicos "vo", "vd" y "va" carecen para el genio de significación específica e intrínseca. Podrían igualmente llamarse "almendra", "banana" y "cereza". Sus significados están definidos *operativamente*, por las posiciones que ocupan en las diversas partes de la definición de la función. Así pues, habría sido un calamitoso error haber escrito, por ejemplo, "(traslatorre (sub 1 n) vo vd va)" al traducir a Lisp el paso 1, porque no contiene indicación de que va y vd hayan de intercambiar sus papeles en dicho paso.

Sigue pendiente una cuestión importante. ¿Qué sucederá cuando nuestro complaciente geniecillo llegue a un paso que diga "traslaundisco"? ¿Saldrá disparado hacia el templo de Benarés y alzará literalmente un disco de oro macizo? O tal vez, más prosaicamente, ¿tomará de un tablero un disco de plástico y lo trasladará de una varilla de plástico a otra? Dicho de otra forma, ¿será posible que en lugar de un mero proceso de cómputo se lleve a cabo algún tipo de acción mecánica?

Bueno, en teoría ello sería perfectamente posible. En efecto, la ejecución de una orden dada en Lisp pudiera conseguir que un brazo mecánico se moviera realmente hasta una localización determinada, alzase cualquier cosa que su brazo mecánico pudiese asir, llevase tal objeto hasta otro emplazamiento específico, y llegado allí lo soltase. En nuestros días, siendo los robots industriales cosa corriente, ello no tendría nada de fantástico. Pero careciendo de brazos y manos mecánicas con los que trasladar objetos materiales, ¿qué significado puede tener "trasladar un disco"?

Una posibilidad obvia y cercanamente emparentada es que sobre una pantalla estuviese representado gráficamente el rompecabezas y que, entonces, la instrucción "traslaundisco" hiciese que un *dibujo* de una mano *pareciera* asir un disco, lo alzase y lo volviese a colocar en algún otro lugar. Tendríamos así una simulación gráfica del pasatiempo, que podría mostrar diversos grados de realismo.

Supongamos, empero, que se carezca del material o de los programas necesarios para representaciones gráficas. Supongamos que todo cuanto queramos sea crear una recta impresa que explique cómo debemos nosotros mover nuestras orgánicas y blandas manos humanas para resolver el problema. Así, en el caso del rompecabezas con tres discos, la receta requerida podría decir "ab ac bc ab ca cb ab", o igualmente bien, "12 13 23 12 31 32 12". ¿Qué podría ayudarnos a lograr esta humilde meta?

Si dispusiéramos de un programa capaz de gobernar el movimiento de un brazo estaríamos interesados no por el valor que nos retornase, sino por la pauta secuencial de "efectos secundarios" que fuera llevando a cabo. Aquí, en cambio, sólo estamos interesados por el valor que nuestro programa va a retornar, a saber, una sucesión de átomos que exhibe ciertas pautas. La lista correspondiente a n = 3 ha de ser construida a partir de dos listas, correspondientes a n = 2. Esta idea fue ya sugerida hacia el final del artículo del mes pasado:

ab ac bc   ab   ca cb ab

En Lisp, para separar grupos, en lugar de intercalar más o menos espacios en blanco se usan paréntesis. Por tanto, nuestro objetivo en el caso n = 3 sería producir una lista "bocata", a saber, "((ab ac bc) ab (ca cb ab))". Un procedimiento para generar una lista a partir de sus componentes consiste en utilizar "cons" reiteradamente. Así pues, si los valores de "almendra" "banana" y "cereza" son respectivamente 1, 2 y 3, el valor retornado por "(cons almendra (cons banana (cons cereza nil)))" será la lista "(1 2 3)". Hay, no obstante, un procedimiento más breve para obtener igual resultado, a saber, escribir "(lista almendra banana cereza)", que retorna el mismo valor. Análogamente, si los átomos "vo" y "vd" están respectivamente ligados con "a" y "b", la ejecución de la orden "(lista vo vd)" producirá "(a b)". La función "lista" es excepcional, pues acepta un número arbitrario de argumentos, e incluso ninguno, en cuyo caso retorna el valor de nil.

Afrontemos el problema de cuál será el valor necesario para que retorne la función "translaundisco". Esta función tiene dos parámetros, que representan varillas, e idealmente desearíamos que retornase un único átomo formado a partir de los nombres de estas varillas, como puede ser "ab" o "12". Por el momento será preferible suponer que los nombres de las varillas son los números 1, 2 y 3. En este caso, para formar el número 12 a partir de 1 y de 2 basta un poco de aritmética: se multiplica el primer número por 10 y se suma el segundo. He aquí las instrucciones necesarias:

```
(def traslaundisco
 (lambda (vo vd)
 (suma (producto 10 vo) vd)))
```

Por otra parte, si los nombres de las varillas hubieran sido literales en vez de numéricos, podríamos haber utilizado una función normalizada en Lisp, llamada "concat", que toma los valores de sus argumentos (cualquiera que sea su número, como en "lista") y los concatena, formando con todos ellos un gran átomo. Por tanto, "(concat 'con 'cat 'enar) retorna el átomo "concatenar". En vista de ello podemos escribir:

```
(def traslaundisco
 (lambda (vo vd) (concat vo vd)))
```

De uno u otro modo hemos resuelto la mitad correspondiente a la línea de salida del problema de trasladar la torre.

La otra mitad del problema consiste en saber qué nos retornará la parte recursiva de "traslatorre". Bueno, eso es bastante fácil. Queremos, sencillamente, que nos retorne una lista "bocata" en la cual dos de las llamadas recursivas flanqueen la llamada única a la operación "traslaundisco". Y así podemos modificar muy ligeramente nuestra definición recursiva anterior, añadiendo la palabra "lista":

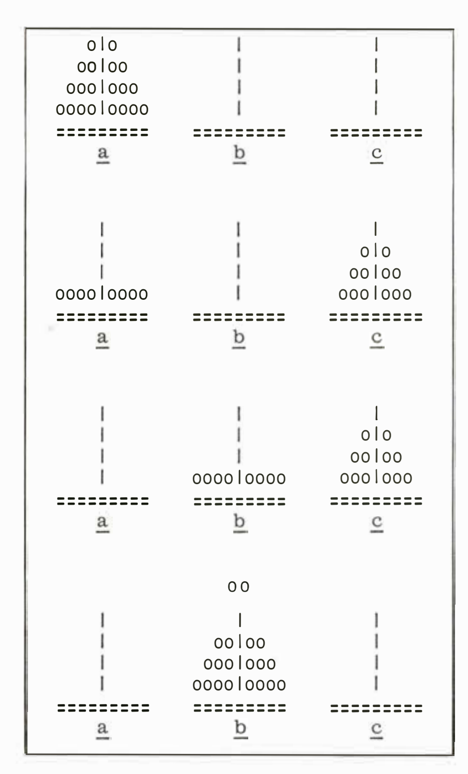
```
(def traslatorre
 (lambda (n vo vd va)
  (cond
   ((eq n 1) (traslaundisco vo vd)
   (t
    (lista
     (traslatorre (sub 1 n) vo va vd)
     (traslaundisco vo vd)
     (traslatorre (sub 1 n) va vd vo))))))
```

Entablemos un breve diálogo con el Lisp-genio:

```
—>  (traslatorre 4 'a 'b 'c)
(((ac ab cb) ac (ba bc ac))
              ab
 ((cb ca ba) cb (ac ab cb)))
—>
```

¡Apabullante! ¡Funciona de veras! ¿No es una preciosidad? En el artículo del mes pasado, a esta lista se la llamó "brahma".

Supongamos que nos hiciera falta suprimir todos los paréntesis interiores, con lo que se imprimiría solamente una larga lista de átomos. Obtendríamos, por ejemplo, "(ac ab cb ac ba bc ac ab cb ca ba cb ac ab cb)" en lugar de la intrincada "brahma". Aunque ligeramente menos informativa, su aspecto es también menos aterrador. En tal caso no nos conviene usar la función "lista" para construir nuestro bocadillo de tres valores, sino que tendríamos que



1. Torre de Brahma de cuatro discos

usar alguna otra función que fuera suprimiendo los paréntesis de los dos valores recursivos situados en los flancos.

Es este un caso donde la función normalizada "anexar" puede echarnos una mano, pues va pegando entre sí cualquier número de listas, despojándolas conforme va haciéndolo de sus paréntesis más externos. Así pues, "(anexa '(a (b)) '(c) nil '(d e))" genera una lista formada por cinco elementos "(a (b) c d e)" en lugar de la lista de cuatro elementos "((a (b)) (c) nil (d e))" que hubiera sido la generada si en la posición funcional hubiera figurado "lista" en lugar de "anexa". Usando "anexar" más una versión ligeramente modificada de traslaundisco podemos formular una versión definitiva de traslatorre que haga lo que pretendemos:

```
(def traslatorre
 (lambda (n vo vd va)
  (cond
   ((eq n 1) (traslaundisco vo vd))
   (t
    (anexa
     (traslatorre (sub 1 n) vo va vd)
     (traslaundisco vo vd)
     (traslatorre (sub 1 n) va vd vo))))))

(def traslaundisco
 (lambda (vo vd) (lista (concat vo
 vd))))
```

Para poner a prueba este programa le pedí al Lisp-genio que resolviese una torre de Brahma de nueve pisos. Lo que casi instantáneamente me retornó puede verse en la figura 2.

Acabamos de analizar un ejemplo de recurrencia bastante complejo, que nos ha hecho devanarnos los sesos no poco. Echemos ahora un vistazo a otra recurrencia que plantea un reto de distinta naturaleza. Tal recurrencia se suscita en una observación casual que hice en mi último artículo. Usé entonces la variable -de exótico nombre- "aysa", mencionando que es un acrónimo recursivo de "aysa (y solamente aysa)". Gracias a este hecho podemos prolongar "aysa" cualquier número de veces. El único principio es que cada aparición de "aysa" a un nivel dado ha de ser reemplazada por la frase "aysa (y solamente aysa)" cuando se vaya a formar el nivel siguiente. La figura 3 nos muestra el proceso. El reto que se nos plantea es el de redactar en Lisp una función que nos retorne el valor "aysa" tras n expansiones recursivas, para algún valor de n. Es irrelevante que para valores de n mucho mayores que 3 la cosa adquiera proporciones ridículamente grandes. ¡Ahora somos teóricos!

Tan sólo hay un problema: cualquier función Lisp ha de retornar como valor una única estructura Lisp (un átomo, o una lista), y las entradas de nuestra tabla no satisfacen tal criterio. Por ejemplo, en el caso n = 2 la entrada consiste en un átomo seguido por dos listas. Para arreglarlo podemos transformar cada entrada de la tabla en una lista, encerrándola entre paréntesis que la abarquen de extremo a extremo. Ahora nuestro objetivo es coherente con las reglas del Lisp. ¿Cómo lo alcanzaremos?

Al pensar recursivamente nos damos cuenta de que la línea de base, o sea, el caso embrionario, se presenta cuando n es igual a 0, y que en los restantes casos la n-ésima línea se construye a partir de la precedente reemplazando el átomo "aysa", cada vez que aparezca, por la lista "(aysa (y solamente aysa))", si bien desprovista de sus paréntesis exteriores. Podemos expresarlo así:

```
(def expansión-de-aysa
 (lambda (n)
  (cond
   ((eq n 0) '(aysa))
   (t
    (sustituye
     'aysa
     '(aysa (y solamente aysa))
     (expansión-de-aysa (sub 1 n)))))))
```

Lo único todavía no especificado es qué significa "sustituir". Hemos de tener el máximo cuidado al definir cómo se ha de efectuar tal sustitución. Consultando cualquier línea de la tabla "aysa" veremos que contiene un elemento más que la línea precedente. ¿Por qué así? Porque el átomo "aysa" se va sustituyendo cada vez por una lista de dos elementos, de cuyos paréntesis, como ya señalé antes, es despojada en el acto de la sustitución. Y esta supresión de paréntesis es la causante de nuestras dificultades. Un ejemplo menos difícil de supresión de paréntesis que la implícitamente definida en "aysa" podría ser "(sustituye 'a '(1 2 3) '(a b a))", cuyo valor debería resultar "(1 2 3 b 1 2 3)". Más que la sustitución exacta de un átomo por una lista, este tipo de sustituciones requiere ir empalmando o anexando en el interior de una lista larga.

Veamos de expresar en Lisp -por recurrencia, como de costumbre- qué significa "sustituir" todas las apariciones del átomo "atm" por una lista, llamada "lysta" en el seno de una larga lista llamada "listalarga". Si quiere pasar un rato entretenido este rompecabezas no le irá mal. Le daré una pista: vea cómo se construye la solución correspondiente al argumento "(a b a)" a partir de la solución correspondiente al argumento "(b a)". Analice igualmente otros casos semejantes, descendiendo hacia el caso embrionario.

El caso embrionario se da cuando listalarga es nil. Nada ocurre entonces, así que la respuesta debería ser nil.

El caso recursivo requiere construir una solución más compleja a partir de una sencilla hipotéticamente dada. Podemos para ello retomar nuestro ejemplo "(a b a)". Construimos la solución compleja "(1 2 3 b 1 2 3)" a partir de la más sencilla "(1 2 3)" agregándole "(1 2 3)". Aceptemos, por un lado, que el propio "(b 1 2 3)" sea una respuesta compleja construida a partir de la respuesta, más sencilla, "(1 2 3)" consando *b* sobre ella. ¿Por qué una requiere consar y la otra anexar? Sencillamente, porque el primer caso contiene el átomo "a", que *sí* es reemplazado, mientras en el segundo, el átomo "b" *no* lo es. Tal observación nos permite intentar la redacción de una definición recursiva de "sustituir", a saber,

```
(def sustituye
 (lambda (atm lysta listalarga)
  (cond
   ((nulidad listalarga) nil)
   ((eq (car listalarga) atm)
    (anexa
     lysta
```

```
 (sustituye atm lysta (cdr lista-
larga))))
  (t
  (cons
   (car listalarga)
   (sustitye atm lysta (cdr lista-
larga))))))
```

Como podemos ver, hay un caso embrionario (donde listalarga es igual a nil), después un caso recursivo caracterizado por "anexar" más otro caso recursivo caracterizado por "consar". Ensayemos esta definición sobre un nuevo ejemplo:

```
—> (sustituye 'a '(1 2 3) '(a (a) b a))
(1 2 3 (a) b 1 2 3)
—>
```

¡Lástima! ¡Estuvo a punto de funcionar! Pero una de las apariciones de "a" ha sido totalmente olvidada. Significa esto que en nuestra definición de "sustituir" se nos pasó por alto alguna eventualidad. En efecto, repasándola vemos que por delante de las narices se nos ha colado una hipótesis cuyo cumplimiento no está garantizado, a saber, que los elementos de listalarga son siempre átomos. Hemos olvidado la posibidad de que listalarga puede contener sublistas. ¿Qué hacer en casos así? Respuesta: efectuar también la sustitución en el seno de tales sublistas. Y también en el interior de las sublistas de sublistas –y así sucesivamente. ¿Se le ocurre algún procedimiento para remendar la definición?

Hemos encontrado ya una recurrencia donde era preciso explorar todos los recovecos de todos los niveles de una estructura: fue la función "recatom" (recuento atómico), vista el mes pasado, en la cual hicimos recurrencia simultáneamente sobre "car" y "cdr". La línea recursiva rezaba "(suma (recatom (car s)) (recatom (cdr s)))". Nuestro proceder habrá de ser aquí muy análogo. Tendremos una línea recursiva caracterizada por *dos* llamadas a la función "sustituir", una para la car de largalista, y otra para la cdr, en lugar de una única para la cdr. Cuando se piensa un poco en ello se comprende que tiene perfecto sentido. Supongamos que fuera preciso reemplazar todos los unicornios de Europa por cuerpoespines. Una forma de alcanzar tan nefando objetivo sería dividir Europa en dos piezas: Portugal (la car de Europa) y el resto (la cdr de Europa). Tras sustituir por cuerpoespines los unicornios de Portugal y hacer otro tanto con los unicornios del resto de Europa, recombinaríamos finalmente las dos piezas, para formar una Europa reunificada. (Es de suponer que ello sugiere una operación de "consar".) Evidentemente, para llevar a cabo tan vil operación sería necesario efectuar una escisión y reunificación análoga en Portugal –y así sucesivamente. Sugiere esto que la etapa recursiva tendrá este aspecto:

```
(cons (sustituye 'unicornio
                 '(cuerpoespín)
                 (car unidad-
                 geográfica))
      (sustituye 'unicornio
                 '(cuerpoespín)
                 (cdr unidad-
                 geográfica)))
```

O sea, con mayor elegancia y generalidad:

```
(cons (sustituye atm lysta (car lista-
      larga))
      (sustituye atm lysta (cdr lista-
      larga)))
```

La línea encabezada "cons" permite así atender tanto el caso en que listalarga es no-atómica como el caso en que sí sea atómica aunque no igual a atm. Para lograr que funcione es preciso aumentar ligeramente el caso embrionario: convendremos en que cuando listalarga no sea una lista, sino un átomo, "sustituye" no produzca efecto alguno. Lo cual es muy conveniente, porque así queda englobada la antigua línea "nulidad", y podemos suprimirla. Al ensamblar todas estas piezas, conseguimos una definición nueva y mejorada:

```
(def sustituye
 (lambda (atm lysta listalarga)
 (cond
  ((átomo listalarga) listalarga)
  ((eq (car listalarga) atm)
  (anexa
   lysta
   (sustituye atm lysta (cdr lista-
larga))))
  (t
  (cons
   (sustituye atm lysta (car listalarga))
   (sustituye atm lysta (cdr lista-
larga))))))))
```

Ahora, cuando pidamos al Lisp-genio "(expansión-de-aysa 2)", éste hará imprimir la lista "(aysa (y solamente aysa) (y solamente (aysa (y solamente aysa))))".

Bien, bien. ¿No ha sido una magnífica proeza? Y si nos pareciera que no es para tanto, siempre podríamos intentar llevarla un paso más allá. Un acrónimo recursivo –que contenga una letra representativa del propio acrónimo– puede ser divertido, pero ¿qué le parecerían acrónimos *mutuamente* recursivos? Podría esto significar, por ejemplo, dos acrónimos en donde cada uno de ellos contuviera una letra que denotase al otro? Un ejemplo podría ser "NOODLES" (fideos) y "LINGUINI" (una especie de macarrones planos y delgados), que significan respectivamente:

"NOODLES (oodles of delicious
    LINGUINI), elegantly served"

"luscious itty-bitty NODDLES got
    usually in Naples, Italy"

De paso, observemos que no sólo es NOODLES indirectamente recursivo, sino que lo es también directamente. Nada de malo hay en ello.

En general, la noción de recurrencia mutua hace pensar en un sistema de estructuras mutuamente entretejidas, cada una de las cuales se define por medio de uno o varios miembros del sistema (incluido, posiblemente, él mismo). Y si estamos aludiendo a una familia de acrónimos mutuamente recursivos, ello significa una colección de palabras en cualquiera de las cuales hay letras que pueden representar a cualquier palabra de la familia.

He de admitir que la noción específica de acrónimos mutuamente recursivos no tiene interés práctico. Es empero muy útil como ejemplo ameno de un fenómeno muy común. ¿Quién no ha reflexionado alguna vez sobre la inevitable circularidad de las definiciones de diccionario? Todos podemos comprender que toda palabra quedará en último extremo definida a partir de un grupo fundamental, que ya no es posible reducir ulteriormente, y que no hace sino dar indefinidamente vueltas en torno a sí. El lector puede entretenerse mirando en un diccionario la definición de alguna palabra corriente, y sustituyendo a su vez por sus definiciones las palabras principales que la definición contenga. Yo llevé a cabo el proceso con la palabra "amor", que era definida como "fuerte afecto, devoción o vínculo hacia una o varias personas". Tras sustituir por sus definiciones las palabras "fuerte", "afecto", "devoción", "vínculo" y "persona", preparé el siguiente potingue: "Un estado o tendencia mental moralmente poderoso, que tiene fuerza de carácter, o voluntad para, o consideración afectuosa,

o lealtad, fidelidad o profundo afecto hacia, uno o varios seres humanos, especialmente en cuanto los distingue de las cosas o los animales inferiores".

Todavía no satisfecho, llevé un paso más allá el proceso entero. He aquí mi resultado: "Un conjunto de circunstancias o atributos que caracterizan a una persona o cosa en un momento dado en, con, o por, lo consciente o lo inconsciente, conjuntamente como una unidad plena de, o que tiene una capacidad o habilidad específica en una manera relativa a, o que trata de, o capaz de hacer la distinción entre el bien y el mal en las conductas, o una inclinación a moverse o actuar en una determinada dirección o manera, que tiene la cualidad o estado de ser fuerte en fortaleza moral, autodisciplina, o reciedumbre de carácter, o el acto o proceso volitivo para, o la consideración, atención o cuidado lleno de ternura por, o la cualidad, estado o instancia de ser fiel a, aquellas personas o ideales que uno se encuentra en obligación de defender o sostener, o la condición, cualidad o estado de ser digno de confianza, o un sentimiento profunda y fuertemente sentido de ternura o cariño hacia una o varias criaturas de, o características de, persona o personas, que vive o existe, o se supone que así lo hace, particularmente en tanto que separado o señalado por diferencias de aquello de lo que fue concebido, expresado, o referido como existente en entidad individual, o de cualquier organismo viviente inferior en rango, dignidad o autoridad, típicamente capaz de moverse por sí mismo, pero no de crear su propio alimento por fotosíntesis".

¿Verdad que es romántico? Qué duda cabe de que el amor es ahora aún más misterioso. En su lúcido y clásico trabajo sobre semántica, *The Tyranny of Words*, Stuart Chase hace un ejercicio similar con la palabra "mente" y muestra igualmente bien cuán opaca es. Evidentemente, en este vórtice de confusión acaban arrastradas tanto las palabras abstractas como las concretas. Mi ejemplo favorito lo descubrí hace muchos años, consultando un diccionario de francés. Definía allí el verbo *clocher* (renquear) como *marcher en boitant* (aproximadamente, "caminar cojeando"), mientras que *boiter* (cojear) era *clocher en marchant* ("renquear al caminar"). El ansioso aprendiz de francés que era yo por entonces recibió una ayuda preciosa, por lo mínima, al consultar este par de definiciones.

Pero ya es hora de retornar a mis acrónimos recursivos. He tenido que dedicar no pocos esfuerzos a preparar una familia de ellos que, para sorpresa mía, tratan preferente, ya que no exclusivamente, de platos de cocina italiana. Todo empezó al ocurrírseme este significado de "tomates":

TOMATOES on MACARONI
(and TOMATOES only),
exquisitely SPICED

Las palabras que dentro de cada frase van en mayúsculas son también acrónimos. He aquí el resto de mi familia mutuamente recursiva:

MACARONI:
MACARONI and CHEESE
(a REPAST of Naples, Italy)

REPAST:
rather extraordinary PASTA
and SAUCE, typical

CHEESE:
cheddar, havarti, Emmentaler
(particularly SHARP Emmentaler)

SHARP:
strong, hearty and rather pungent

SPICED:
sweetly pickled in CHEESE
ENDIVE dressing

ENDIVE:
egg NOODLES, dipped
in vinegar eggnog

NOODLES:
NOODLES (oodles of delicious
LINGUINI), elegantly served

LINGUINI:
LAMB CHOPS
(including NOODLES),
got usually in northern Italy

PASTA:
PASTA and SAUCE (that's
ALL!)

ALL!:
a luscious lunch

SAUCE:
SHAD and unusual COFFEE
(eccellente!)

SHAD:
SPAGHETTI, heated al dente

SPAGHETTI:
standard PASTA, always good,
hot
particularly (twist, then ingest)

COFFEE:
choice of fine flavors,
particularly ESPRESSO

ESPRESSO:
excellent, strong, powerful, rich
ESPRESSO, suppressing sleep
outrageously

BASTA!:
belly all stuffed (tummy ache!)

LAMB CHOPS:
LASAGNE and meatballs,
casually heaped onto
PASTA SAUCE

LASAGNE:
LINGUINI and SAUCE
and GARLIC

```
(ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ab ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca bc
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ab ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ca bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ab ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca bc
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ca bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca bc
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ab ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca bc
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ab ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ca bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ab ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ca bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca bc
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ca bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ab ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca bc
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ab ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ca bc ba ca bc ab ac bc ba ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab ac bc ba ca bc ab ac bc ab ca cb ab ca bc ba ca cb
 ab ac bc ab ca cb ab)
```

2. El Lisp-genio nos da las jugadas que resuelven la torre de altura 9

(NOODLES everywhere!)

RHUBARB:
RAVIOLI, heated under butter
and RHUBARB (BASTA!)

RAVIOLI:
RIGATONI and vongole in oil,
lavishly introduced

RIGATONI:
rich Italian GNOCCHI
and TOMATOES
(or NOODLES instead)

GNOCCHI:
GARLIC NOODLES
over crisp CHEESE,
heated immediately

```
n = 0:   aysa

n = 1:   aysa
          (y solamente aysa)

n = 2:   aysa
          (y solamente aysa)
           (y solamente aysa (y solamente aysa))

n = 3:   aysa
          (y solamente aysa)
           (y solamente aysa (y solamentea aysa))
             (y solamente aysa (y solamente aysa) (y solamente aysa (y solamente aysa)))
```

3. Breve tabla de ampliaciones recurrentes de la variable denominada "aysa"

Cualquier gourmet puede advertir que no se ha intentado seriamente que cada término quede definido por su correspondiente frase, sino que está relacionado más o menos caprichosamente con ella.

¿Qué sucede al empezar a expandir alguna palabra como, por ejemplo, "pasta"? Al principio obtenemos sencillamente "PASTA and SAUCE (that's ALL!)". La etapa siguiente produce "PASTA and SAUCE (that's ALL!) and SHAD an unusual COFFEE (eccellente!) (that's a luscious lunch)". Podríamos, como es obvio, ir desarrollando acrónimos indefinidamente o, al menos, hasta llenar el universo hasta los bordes con boquilicuefacientes descripciones de platos italianos. Pero, ¿y si siendo menos ambiciosos pretendiéramos llenar solamente media página con tales descripciones? ¿Cómo podríamos dar con algo que detuviera a medio camino esta recurrencia aparentemente desprovista de "línea de base"?

La frase clave es precisamente ésta, y la solución que conlleva es: introducir un mecanismo que permita a la recurrencia salir "por la planta baja". Hasta ahora eso no ocurría porque a cada acrónimo le estaba permitido en cada etapa desarrollarse, es decir, generar nuevos acrónimos. Pero, ¿qué sucedería si en lugar de eso mantuviéramos estricto control del proceso de proliferación, mostrándonos magnánimos en las "generaciones" iniciales, mas permitiendo sólo luego a un número progresivamente menor de acrónimos que se multiplicasen conforme van sucediéndose las progenies? Sería a modo de un secoya, que comienza por una sola "rama" (el tronco); luego esa rama se multiplica y genera "descendencia", a saber, una primera generación de ramas secundarias, que a su vez engendran una progenie mayor, aunque en último extremo se produzca una "salida por la base" como resultado de que las ramitas tiernas no pueden ramificarse ya. (Parece como si en el curso de la evolución a los árboles se les hubieran cruzado las conexiones, porque, para ellos, "la salida por la base" se produce generalmente ¡en lo alto de sus copas!)

Si este proceso fuese completamente regular todos los secoyas tendrían el mismo aspecto, dándole así la razón al memorable aforismo del ex-gobernador Reagan: "Visto un secoya, vistos todos". Desafortunadamente, los secoyas (y presumiblemente otras cosas) son más complejos de lo que al gobernador Reagan se le alcanzaba, y tenemos que aprender a habérnoslas con una multitud de cosas denotadas por un mismo nombre; multitud y variedad que es consecuencia de lo aleatorio de los puntos de ramificación, del ángulo en que crecerán las nuevas ramas, de cual será su tamaño, etcétera.

Observaciones análogas son válidas para los "árboles" de acrónimos mutuamente recursivos. Si al desarrollar "tomates" hubiéramos tomado siempre las mismas decisiones acerca de qué acrónimos desarrollar y cuándo, entonces habría un solo tipo de desarrollo "rhubarb"; y así tendría sentido una afirmación del siguiente tenor: "Si Vd. ha visto un 'rhubarb', los habrá visto todos". En cambio, admitiendo cierto grado de azar para decidir qué y cuándo desarrollar, obtendremos muchas variantes, que mostrarán todas cierta semejanza característica, aunque mucho más esquiva de precisar.

¿Cómo podremos conseguirlo? La noción ideal a tener en cuenta es la de generador de números aleatorios, que hacen en computación el papel equivalente al de lanzar dados o monedas. Haremos que la decisión de desarrollar o no un acrónimo dado dependa del resultado de tales lanzamientos virtuales de monedas. En las primeras etapas del desarrollo prepararemos las cosas para que sea muy probable que la moneda salga "cara" (desarrollar); en etapas posteriores habrá de ser progresivamente más probable que salgan cruces (no desarrollar). Nos valdremos para ello de la función Lisp llamada "rand". Esta función no tiene argumentos, y cada vez que es nombrada nos retorna un número real comprendido entre 0 y 1, de valor impredictible. (En realidad, esto es exagerado. Si se sabe cómo se calcula, tal valor es predictible ciento por ciento; pero como el algoritmo es bastante complicado, a casi todos los observadores, y para casi todos los propósitos, el comportamiento de la función "rand" resulta tan errático que puede tomarse como enteramente aleatorio. La historia de la generación de números aleatorios es fascinante, digna de un artículo para sí sola.)

Si queremos que un suceso acontezca con probabilidad del 60 por ciento, empezaremos pidiendo un valor de rand. Cuando tal valor resulte ser 0,6 o menos, seguimos adelante; en caso contrario, no. Puesto que tomado un lapso de tiempo grande, la función rand rocía uniformemente sus valores sobre el intervalo comprendido entre 0 y 1, tendremos la señal de continuar en el 60 por ciento de las veces.

Y basta ya de decisiones aleatorias. ¿Cómo haremos para que el acrónimo se expanda cuando se le ordene así? No es demasiado difícil. Supongamos que cada acrónimo sea una función en Lisp, como en el siguiente ejemplo:

```
(def tomatoes
 (lambda ()
  '(tomatoes on macaroni
   (and tomatoes only)
   exquisitely spiced))))
```

La función "tomatoes" no requiere argumentos, y retorna, sencillamente, la

lista de palabras correspondiente a su desarrollo. Nada más sencillo.

Supongamos ahora que tenga una variable llamada "acrónimo" cuyo valor sea cierto acrónimo particular, aunque no sabemos cuál. ¿Cómo hacer que tal acrónimo se desarrollase? En la forma en que lo hemos preparado, el acrónimo debe actuar como apelación a una función. Para que un átomo pueda invocar una función ha de ser la car de una lista,como en los ejemplos "(suma 2 2)", "(rand)" y "(rhubarb)". Si sencillamente escribiéramos "(acrónimo)", el átomo literal "acrónimo" sería tomado por el Lisp-genio como nombre de una función. Ello sería, empero, un malentendido, pues no es el átomo "acrónimo" quien ha de servirnos como nombre de función, sino su *valor*, ya sea éste "macaroni", "cheese" o el que fuere.

Para conseguirlo recurriremos a un pequeño truco. Si el valor del átomo "acrónimo" es "rhubarb", y si yo le pido al Lisp-genio "(lista acrónimo)", el valor que me retornará será la lista "(rhubarb)". Sin embargo, el Lisp-genio no considerará esto una orden mía, que me gustaría ver ejecutada, sino como un fragmento inerte de Lisp. No puede leer mi pensamiento. ¿Cómo conseguiré que realice la operación deseada? La solución está en una función que ahora recuerdo, llamada "evaluar", que hace que el genio considere cualquier estructura de datos como un deseo a ejecutar. En este caso me basta decir "(eval (lista acrónimo))" para obtener la lista "(ravioli, heated under butter and rhubarb (basta!))". Y si el acrónimo tuviera un valor distinto, el genio me facilitaría una lista diferente.

Tenemos ahora justamente las ideas suficientes para construir una función capaz de desarrollar acrónimos mutuamente recursivos, creando frases largas, aunque finitas, cuyos tamaños y estructuras estén determinados por muchos "lanzamientos" virtuales de la moneda "rand". En lugar de ir llevando al lector paso a paso hacia la construcción de esta función, me limitaré a presentársela y someterla a su atención. Tiene por modelo cercano la función "sustituir", construida hace muy poco:

```
(def desarrollar
 (lambda (probabilidad frase)
  (cond
   ((átomo frase) frase)
   ((es-acrónimo (car frase))
    (cond
     ((menor-que (rand) probabilidad)
      (anexar
       (desarrollar (eval (lista (car frase)))
                    (reducir probabilidad))
       (desarrollar (cdr frase) probabi-
        lidad)))
     (t
      (cons (car frase)
            (desarrollar (cdr frase)
                         probabi-
                         lidad)))))
   (t
    (cons
     (desarrollar (car frase)
                  (reducir probabilidad))
     (desarrollar (cdr frase) probabi-
      lidad))))))
```

Observemos que "desarrollar" tiene dos parámetros. Uno de ellos representa la frase a desplegar, mientras que el otro representa la probabilidad de expansión de cualesquiera acrónimos del nivel principal de la frase dada. (Así pues, el valor del átomo "probabilidad" será siempre un número real comprendido entre 0 y 1.) La probabilidad de expansión ha de ir decreciendo conforme las llamadas van haciéndose más y más recursivas. Tal es el motivo de que las líneas que piden la expansión de (car frase) lo hagan con probabilidad más reducida. Para ser exactos, podemos definir, por ejemplo, que la función "reducir" sea como sigue:

```
(def reducir (lambda (x) (producto x
0,8)))
```

Así, cada vez que se desarrolle el acrónimo, la probabilidad de que su progenie se expanda será solamente 0,8 veces de la que era, lo que implica que los acrónimos más profundamente situados tienen probabilidad de proliferación tendente a cero. Puede usarse cualquier otro coeficiente de reducción; el valor 0,8 nada tiene de sagrado, excepto porque parece proporcionarme buenos resultados.

"Es-acrónimo" constituye la única de las funciones que figuran en la definición anterior y que no ha sido descrita todavía. Su nombre es autodescriptivo. La función prueba primero si su argumento es o no un átomo; si no lo es, retorna nil. Cuando el argumento sí es un átomo, la función procede a ver si tal átomo está o no definido como nombre de una función, y en particular, definido como acrónimo. Cuando el átomo tiene una tal definición, es-acrónimo retorna el valor t; en los restantes casos retorna nil. La forma exacta en que esto se realiza depende de la variedad específica de Lisp que utilicemos; por cuyo motivo no la muestro aquí explícitamente. Con Franz Lisp sólo se requiere una línea.

Quizás haya usted notado que hay dos cláusulas condicionales estrechamente próximas que comienzan por "t". ¿Cómo es que un "en otro caso" va pisando tan de cerca los talones del otro? La verdad es que pertenecen a condicionales distintas, encajadas una en otra. La primera "t" (que corresponde a la condicional más interior) se aplica a casos donde sabemos estar manejando un acrónimo, pero donde nuestra moneda de la suerte, en lugar de salirnos cara, nos ha salido cruz (lo que equivale a decidir no desarrollar); la segunda "t" (perteneciente al condicional exterior) se aplica en el caso de que hayamos descubierto que lo que tenemos entre manos no es un acrónimo.

La función "expansión" saca partido de una de las más potentes características del Lisp, que es la capacidad de este lenguaje para tomar las estructuras de datos que ha creado y darles tratamiento de elementos de código (es decir, presentárselas al Lisp-genio con carácter de órdenes). Solamente lo hemos hecho aquí en el nivel más rudimentario. Hemos envuelto átomos entre paréntesis, y la minúscula lista resultante, evaluada. El trabajo preciso para fabricar la estructura de datos ha sido prácticamente nulo en este ejemplo; pero en otros casos puede que hayan de "consarse" complejas secciones estructurales, que luego serán presentadas al Lisp-genio para su evaluación. Tales bloques de código pueden ser definiciones de nuevas funciones, o cualquier número de otras cosas. La idea principal es que en Lisp se tiene la capacidad de "elevar" una estructura de datos, meramente portadora de información, hasta el nivel de "agente animado", en cuyo nivel se convierte a su vez en manipuladora de estructuras inertes. Este ciclo programa-datos, llamado "bucle", puede continuar indefinidamente, al tiempo que las estructuras implicadas en él se expanden, se repliegan y retuercen sobre sí mismas, modificándose o modificando indirectamente a otras.

Ciertos tipos de estructuras inertes portadoras de información son llamadas a veces "conocimientos declarativos", a causa de su forma, que recuerda abstractamente las frases "declarativas", y en razón también de que encodifican hechos relativos al mundo ("conocimientos"), que son accesibles revisando un índice, un poco a la manera de los hechos recogidos en libros y accesibles a los humanos. Por contras-

te, los elementos de código, "activos" o "animados", suelen denominarse "conocimientos procesales", porque definen series de acciones ("procedimientos") que manipulan realmente estructuras de datos, y que son "conocimientos" precisamente en tanto que sistemas de habilidades y destrezas adquiridas por el programa; algo así como la pericia que los humanos llegan a conseguir tras largas sesiones de práctica repetitiva. A veces, estos dos tipos de conocimientos contrastantes son denominados "conocimiento de qué sea la cosa" y "conocimiento de cómo".

Distinción que a los biólogos puede recordarles las diferencias entre genes, que son estructuras relativamente inertes implantadas en el interior de las células, y los enzimas, que son todo menos inertes. Los enzimas son los agentes animados de la célula; transforman y manipulan todas las estructuras inertes en formas increíblemente complejas. Además, los bucles de programa y datos del Lisp pudieran sugerir a los biólogos la forma en que los genes dictan cómo deben ser los enzimas y cómo los enzimas manipulan genes (entre otras muchas cosas). Así pues, el bucle procesal-declarativo, programa-datos, del Lisp nos proporciona un ejemplo rudimentario, aunque muy útil y tangible, de las pautas fundamentales que se encuentran en la base misma de la vida: la capacidad de estructuras pasivas para controlar su propio destino, creando y destruyendo a tal fin estructuras activas cuyas formas dictan.

Hemos venido hablando del Lisp-genio como si se tratara de un agente misterioso, sin preguntar dónde podemos encontrarlo o qué lo hace funcionar. Resulta que una de las más interesantes características del Lisp es que puede describirse cómodamente la naturaleza completa del Lisp-genio en el propio Lisp. Con otras palabras, el intérprete de Lisp puede ser fácilmente redactado en Lisp. Evidentemente, si no hay un intérprete de Lisp que haga funcionar *ese* intérprete de Lisp, tal ejercicio podría parecer tan absurdo como carente de sentido, un poco a la manera de esas floridas introducciones de algunos cursos de idiomas, que explican cómo debe hacerse para mejor aprender el nuevo idioma... en ese mismo idioma. La cosa no es, empero, tan estúpida como se parece estar diciendo.

Pues, en primer lugar, si tenemos ya algunas nociones de ese idioma, el preámbulo podrá servirnos para lograr mayor altura en él por esfuerzo propio, como si dijéramos, a base de tirar hacia arriba de los cordones de los zapatos. En efecto, a partir de cierto punto, las explicaciones acerca del español, por ejemplo, dadas en español pueden resultar sumamente útiles. Y lo que es más, tal punto no se encuentra muy lejos del nivel inicial. Por consiguiente, todo cuanto se precisa es adquirir primero, y por medios propios, un "núcleo"; después ya podemos empezar a elevarnos "tirando de los cordones". Los niños encuentran apasionante cuando, conforme van leyendo, comienzan por sí mismos a aprender frases nuevas por el escueto método de encontrárselas varias veces seguidas. Su vocabulario se agranda entonces a saltos y zancadas. Y lo mismo sucede en cuanto un sistema contiene un Lisp "nuclear"; el resto del intérprete de Lisp puede ya escribirse en tal lenguaje; así es como se hace.

El hecho de que pueda redactarse fácilmente en Lisp el intérprete de Lisp no es mera chiripa, resultante de alguna recóndita característica del Lisp. A causa del bucle programa-datos, el Lisp se presta muy bien para redactar intérpretes de todo tipo de lenguajes de ordenador, lo que significa que el Lisp puede servir de base para la construcción de otros lenguajes.

A fin de expresarlo más vívidamente, imaginemos que haya usted diseñado sobre el papel un nuevo lenguaje, que llamaremos "Nital". Si verdaderamente supiera usted cómo tiene que funcionar su Nital, no tendría excesiva dificultad para redactar en Lisp un intérprete para él. Una vez materializado su intérprete de Nital, éste se convierte esencialmente en un nuevo genio, al que podemos formular deseos (en Nital), que serán retransmitidos al Lisp-genio (en Lisp). Evidentemente, todos los dispositivos que permiten al "metagenio" del Nital hablar con el Lisp-genio se encuentran a cargo del Lisp-genio. ¿Será, pues, el nuevo lenguaje pura apariencia? Al hablar en Nital ¿estaremos haciendo algo más que hablar encubiertamente en Lisp?

Bueno, cuando los diplomáticos americanos dialogan con sus colegas rusos por intermedio de intérprete, ¿están en realidad hablando encubiertamente en ruso? ¿Estará tal vez la clave del problema en saber si la lengua materna de la intérprete fue ruso o inglés, sobre el cual se construyó la otra lengua como idioma secundario? Supongamos que finalmente nos enteramos de que el lenguaje nativo de la intérprete era el lituano, que sólo llegó a aprender inglés de adolescente, y que el ruso lo aprendió en la universidad, por cierto, en clases dadas en inglés. ¿Pensaremos entonces que, cuando hable en ruso, en el fondo está realmente hablando en inglés, o mejor todavía, que realmente habla en lituano doblemente disfrazado?

De manera análoga, podríamos encontrarnos con que el intérprete de Lisp está realmente escrito en Pascal o en algún otro lenguaje de ordenador. Y entonces alguien podría igualmente arrancar "la máscara del Pascal" y mostrarnos que todas las instrucciones están escritas en *lenguaje de máquina*, ¡con lo que estaremos engañándonos por completo a nosotros mismos si creyéramos que la máquina habla en realidad en Nital, Pascal, Lisp, o cualquier otro lenguaje de alto nivel!

Cuando los intérpretes van a caballo unos de otros se plantea siempre la cuestión de qué nivel tomaremos como último para no profundizar más. Por mi parte, raramente me paro a pensar qué subyace al intérprete de Lisp; y así, cuando manejo el sistema Lisp siento estar hablando con "alguien" cuyo "lenguaje nativo" es Lisp. Análogamente, cuando trato con personas, no se me ocurre pensar cuál es la composición de sus cerebros, ni tampoco los imagino como grandes cúmulos de descargas neuronales. A mi sistema perceptivo le resulta natural reconocer a las personas a un determinado nivel, y no por debajo de él.

Si alguien tuviera que redactar un programa capaz de tratar en chino preguntas y respuestas sencillas relativas a visitas a restaurantes, y si tal programa estuviera escrito en el lenguaje imaginario "SIGLAS" (Simulador Inteligente Gastro-Lingüístico Asiático Surorienta1) yo podría optar entre considerar que el sistema habla genuinamente en chino (suponiendo que su pronunciación fuese verosímil, y tuviera la velocidad adecuada) o que su idioma es, genuinamente, SIGLAS. Puedo cambiar a voluntad mi punto de vista. El que termine por adoptar dependerá sobre todo de consideraciones de índole práctica, tales como el tema que en ese momento más me interese (los restaurantes chinos o el funcionamiento de las gramáticas), o de hasta qué punto la velocidad del nivel dado se ajuste a la de mi propio cerebro y, desde luego, de si domino mejor el idioma chino o el lenguaje SIGLAS. Si los ideogramas chinos son para mí sólo pinceladas y garabatos, seguramente optaré por el enfoque SIGLAS; por otra parte, si pa-

ra mí SIGLAS es nada más un rimero de términos técnicos incomprensibles, tal vez me incline por admitir que aquello es chino. Y si descubro además que SIGLAS se ha construido sobre la base de un sistema Lisp, tendré un tercer punto de vista a considerar; y así sucesivamente.

Empero, al apilar intérprete sobre intérprete el sistema pronto pierde su eficacia. Es como hacer funcionar un motor con electricidad producida por una serie de generadores, donde cada generador hubiera de funcionar con energía tomada del paso precedente. El rendimiento sería muy pequeño, pues en cada etapa se pierde mucha potencia. Por lo general no hay necesidad de montar muchos generadores en serie; en cambio, con los intérpretes de lenguajes de alto nivel puede no haber otra solución. Si no se tiene una máquina cuyo lenguaje de máquina sea Lisp, lo que se hace es construir un intérprete de Lisp para la máquina de que se disponga, y se hace funcionar el Lisp de ese modo. Y si deseamos disponer también de Nital o de SIGLAS, se los construye superpuestos a esta "máquina virtual de Lisp". Un proceder tan indirecto puede resultar de una ineficiencia desesperante, y ser causa de que la máquina virtual SIGLAS o Nital sea docenas de veces más lenta de lo que nos gustaría.

Durante estos últimos años se han conseguido importantes logros en la circuitería de los ordenadores, y hoy se dispone de máquinas que funcionan en Lisp a nivel de soporte físico ("hardware"). Significa esto que tales máquinas "hablan" Lisp en un sentido bastante más profundo –"más fluidamente", digamos– que las máquinas virtuales de Lisp. Significa también que cuando se está trabajando en una de tales máquinas se está "nadando" en un medio –un ambiente– Lisp. Es preciso tener en cuenta que un ambiente Lisp conlleva mucho más de lo hasta ahora descrito, por ser más que un mero lenguaje con que redactar programas: hace falta un programa "editor", para crear, modificar y corregir los programas, un programa de "depuración" (*debugging*) con el que detectar y corregir fácilmente los errores de programación, y muchas otras características, diseñadas todas ellas para ser mutuamente compatibles.

Tales máquinas, aunque todavía costosas y hasta cierto punto en fase de desarrollo, están abaratándose rápidamente. Se dispone ya de las manufacturadas por compañías como LMI (LISP Machine, Inc.), Symbolics, Inc., ambas de Cambridge, Massachusetts, y por otras más antiguas, como la Xerox. Existen también versiones de Lisp para la mayoría de microordenadores; basta ojear algunas de las revistas dedicadas a ordenadores personales para encontrarlas anunciadas.

En conclusión, ¿por qué tiene el Lisp tanto predicamento en el campo de la inteligencia artificial? No hay respuesta única ni sencilla. ¿Sería válido decir que el Lisp "es el lenguaje del pensamiento"? Con certeza, no. Hace 10 o 20 años, tal vez la gente dedicada a inteligencia artifical lo hubiera defendido así, pero muy pocos serían hoy tan osados. Una de las razones de que el Lisp haya seguido en boga es, como ya escribí en mi primer artículo, que el Lisp es claro. Tanto así que una función bien definida en Lisp puede serle a un "líspero" tan grata como un soneto a un amante de la poesía. Mi colega Dan Friedman esta componiendo un libro titulado *Lisp Poems*, y por ello acucia a los lectores interesados para que le envíen sus logros más elegantes. Su dirección es Computer Science Department, Indiana University, Bloomington, Indiana 47405 (USA).

¿Qué propiedades le han conferido al Lisp tan distinguido papel en la investigación de inteligencia artificial? Yo diría que la lista siguiente es prácticamente exhaustiva:

1. El Lisp es claro.
2. El Lisp es interactivo.
3. El Lisp está centrado en la noción de lista y de manipulación de listas, y las listas son estructuras de datos sumante versátiles y flexibles.
4. Al tener el código Lisp igual forma que los datos, es fácil de manufacturar y de hacer funcionar.
5. Es fácil construir y experimentar con intérpretes para otros lenguajes, redactándolos en Lisp.
6. A mucha gente le resulta sencilla y natural la "natación" en un ambiente Lisp.
7. El Lisp está empapado de "espíritu de recurrencia".

Tal vez la clave resida en este último enunciado, de naturaleza bastante etérea. Por alguna razón, en inteligencia artificial hay mucha gente que parece sentir allá, en lo más hondo, que la recurrencia, en alguna de sus formas, está engarzada con el "truco" de la inteligencia. Es una corazonada, una intuición, un sentimiento vagamente percibido, una creencia ligeramente mística, que ciertamente yo comparto.

Taller y laboratorio

Cómo construir una lente de hielo y qué acontece cuando reposa el café turco

Jearl Walker

A primera vista, diríase que construir una lente con hielo constituye algo más bien inverosímil. Sin embargo, es perfectamente posible y ofrece varios caminos de experimentación. Se trata de una idea que mereció ya un puesto destacado en una de las narraciones de fantasía científica de Julio Verne, *El Desierto de hielo*. Matthew G. Wheeler, de McBride, Columbia Británica, acaba de convertirla en realidad.

En la novela de Verne un grupo de exploradores sobreviven a una catástrofe en el Artico gracias a una lente de hielo. El capitán del barco había zarpado con el propósito de alcanzar el Polo Norte, pero un motín los había dejado, a él y a unos pocos fieles, con un navío destrozado, a punto de hundirse, y escasas provisiones. Su única posibilidad de supervivencia era llegar a otro barco que se había abandonado al quedar atrapado entre los hielos. Entonces, el grupo obtiene leña de su maltrecha embarcación, empaqueta sus exiguas vituallas, recoge el pedernal y el eslabón para hacer fuego y parte.

Sus víveres comienzan a agotarse. Acechan y cazan un oso polar; pero, mientras estaban ocupados en ello, se apaga el fuego y se pierden el pedernal y el eslabón. La comida que les queda basta para el resto del viaje, pero si no consiguen recuperar el fuego la amenaza de muerte por congelación puede hacerse real.

La óptica les salvará. Un físico integrante del grupo, al observar la luminosidad del Sol, recuerda cómo hacer fuego con una lente y, carente de una de vidrio, decide construir una de hielo. Busca así un lugar donde el hielo aparezca traslúcido, corta un pedazo de unos treinta centímetros de diámetro y lo conforma groseramente con un hacha. Luego lo talla y lo alisa con un cuchillo y lo pule con sus manos desnudas "hasta conseguir una lente de tal transparencia que parecía hecha de cristal noble". Colocaron luego la lente de hielo encima de un poco de yesca, enfocando sobre ésta los rayos solares y, en cuestión de segundos, la yesca prendió. "Las brasas no tardaron en crepitar y no pasaron muchos minutos antes de que el olor de filete de oso a la parilla" despertase a los marineros.

Nuestro físico había construido una lente convexa. De tamaño mayúsculo, recogía una buena cantidad de luz solar, por cuya concentración en el foco no tardó en reunir energía suficiente para encender un fuego.

Este relato lo incluí en una de mis colaboraciones habituales en Quirks and Quarks, programa radiofónico de divulgación científica que emite la Canadian Broadcasting Corporation. Describí también allí cómo hacer una lente convexa de una manera menos complicada utilizando vidrios de reloj curvos. (O sea, las piezas de vidrio que cubren las esferas de los relojes. Entre los útiles de cualquier laboratorio de química corriente pueden encontrarse vidrios similares, de mayor tamaño.) Basta para ello verter agua en dos vidrios de reloj y esperar que se congele. Luego se forma la lente uniendo los dos trozos de hielo.

Wheeler, uno de mis oyentes, decidió mejorar mi idea aprovechando el hielo que se forma en un abrevadero de ganado. De elemento de partida le sirvió un trozo de hielo de unos dos o tres centímetros de grueso y exento de burbujas. Para darle forma lo introdujo en la tapadera de una pasterizadora de leche, cuya curvatura es similar a la de una tapa de cacerola. El conjunto lo puso a flotar en una pila llena de agua caliente, para que el calor fundiera el fondo del hielo hasta que éste se adaptara al fondo de la tapadera. De tanto en tanto hacia girar el hielo para aumentar la simetría de su fondo. Luego le dio la vuelta a la pieza y conformó la otra cara.

Con ello tenía ya una lente de hielo convexa. Para ensayar su poder de concentración (y remedar al físico de la narración de Verne) la sujetó de canto y enfocó la luz del Sol sobre un periódico: "El papel se chamuscó inmediatamente, humeó y rompió a arder". Wheeler me expone que la lente le funcionaba mejor cuando el aire estaba sobre cero, pues a temperaturas inferiores sobre ella se condensan gotas de agua, que alteran su superficie e interfieren en la focalización de la luz.

Valiéndose de la base cóncava de un recipiente presurizado, construyó otra lente más pequeña, que, cuando la colocó delante de su cámara de 35 milímetros, actuó de teleobjetivo de una distancia focal de unos 80 milímetros. Wheeler la ensayó al aire libre durante un día frío y nevado. La lente resultó demasiado rápida para captar escenas de nieve, por cuya razón la diafragmó a $f\,9$ con un iris de cartón. Este lo instaló entre un anillo de tubo de plástico negro y una visera fotográfica de caucho de tipo corriente. El anillo eliminaría la luz que no hubiese atravesado la lente.

Al principio, Wheeler sujetaba la lente con la mano enguantada, pero, posteriormente, descubrió que le era más cómodo emplear una montura de cartón. A ésta aseguró la lente con tres pestañas de cartón engrapadas al cartón, y, delante, colocó una visera también de cartón. Una vez instalada la lente en su montura, apretó el conjunto de cartón y lente contra la visera de caucho y ajustó su posición. Según Wheeler, una lente de hielo trabajará perfectamente durante 15 minutos y funcionará aun cuando la temperatura del aire llegue a los 16 grados Celsius.

Wheeler ha construido un buen número de lentes empleando una amplia variedad de moldes. A veces, parte de agua puesta a congelar. Según afirma, consigue el hielo más traslúcido cuando el agua comienza a solidificarse por arriba; y, a tal fin, aísla los costados y la base del molde.

Si alguien se decide a construir una lente de hielo con un par de vidrios de reloj, que haga la prueba de juntar dos trozos de hielo. Estos deben mantener un buen contacto. Para ello deberá frotarlos uno sobre otro o pasar el dedo para humedecer y alisar las superficies de contacto, pues cuando se suelden entre sí la cohesión debe ser sólida.

Eligiendo los moldes convenientes se construyen lentes de hielo convexas, cóncavas y concavoconvexas. Por ser el índice de refracción del hielo menor que el del vidrio, las lentes de hielo desvían los rayos solares menos que las de vidrio de la misma hechura. De ahí que una lente de hielo (sea convexa o cóncava) tendrá una distancia focal mayor que una lente similar hecha de vidrio.

El café turco es una bebida fuerte que se prepara con azúcar y café finamente molido. El azúcar se requie-

1. Matthew G. Wheeler prendiendo fuego a un periódico con una de sus lentes de hielo

2. Paisaje de Columbia Británica fotografiado con una lente de hielo montada en una cámara

re para neutralizar el gusto amargo del café, pero es opinión mía que además interviene en la elaboración de la infusión. Se prepara ésta calentando una mezcla de café molido, azúcar y agua dentro de un ibrik, que es un recipiente levemente ahusado hacia arriba y provisto de un mango largo. El café rompe a hervir enseguida y se vierte en una tacita pequeña, de un tamaño aproximadamente la mitad de una taza de café con leche. Lo que entra en la taza es una mezcla de líquido, espuma y granos de café. La espuma es sumamente apreciada en los países del Oriente Medio. Hay quien dice que constituye una injuria para un invitado servirle una taza de café sin espuma.

Antes de probarlo, debe dejarse reposar en la taza algunos minutos. Ello se debe a que los granos de café tardan en depositarse en el fondo y, si la bebida se cata antes de tiempo, el líquido resulta desagradablemente pulverulento. Cuando los granos se han posado, el líquido se sorbe poco a poco hasta que sólo queda el poso, cuya consistencia es similar a la de un lodo.

He pasado horas reflexionando acerca de la preparación y sedimentación del café turco mientras preparaba una tacita tras otra. Me parece que es la única modalidad en que las partículas de café quedan en el líquido tras la infusión. Encierra también interés en otros aspectos. ¿Por qué debe emplearse café finamente molido? (Al comprarlo para preparar un café turco, debe pedirse de clase extra-molida.) ¿Por qué se calienta el azúcar junto con el café y no se añade luego como ocurre en otros casos? ¿Por qué el líquido forma repentinamente espuma en los costados del ibrik mientras se infunde el café? ¿Cuánto debe reposar el café en la taza para que al sorberlo esté libre de polvillo?

El café que yo consumo es de tueste por el método francés o vienés y está finamente molido. Ambos son más oscuros que los otros que se venden habitualmente, lo que significa que se tuestan durante más tiempo y que han perdido mayor cantidad de sustancias volátiles. Dudo, no obstante, que el tipo de café afecte a otra cosa que no sea al propio aroma de la bebida.

El mío lo preparo en un ibrik de hojalata de capacidad suficiente para dos tacitas, cada una de las cuales admite unos 60 centímetros cúbicos, aunque acostumbro a preparar una sola tacita por vez. Evito así que la espuma se forme a tal velocidad que se derrame en cascada sobre el fuego. Pongo en el ibrik una cantidad de agua algo inferior a la necesaria para llenar una tacita. Luego añado al agua dos cucharaditas de café molido y de una a cuatro cucharaditas de azúcar. (Una cucharadita equivale a cinco centímetros cúbicos.)

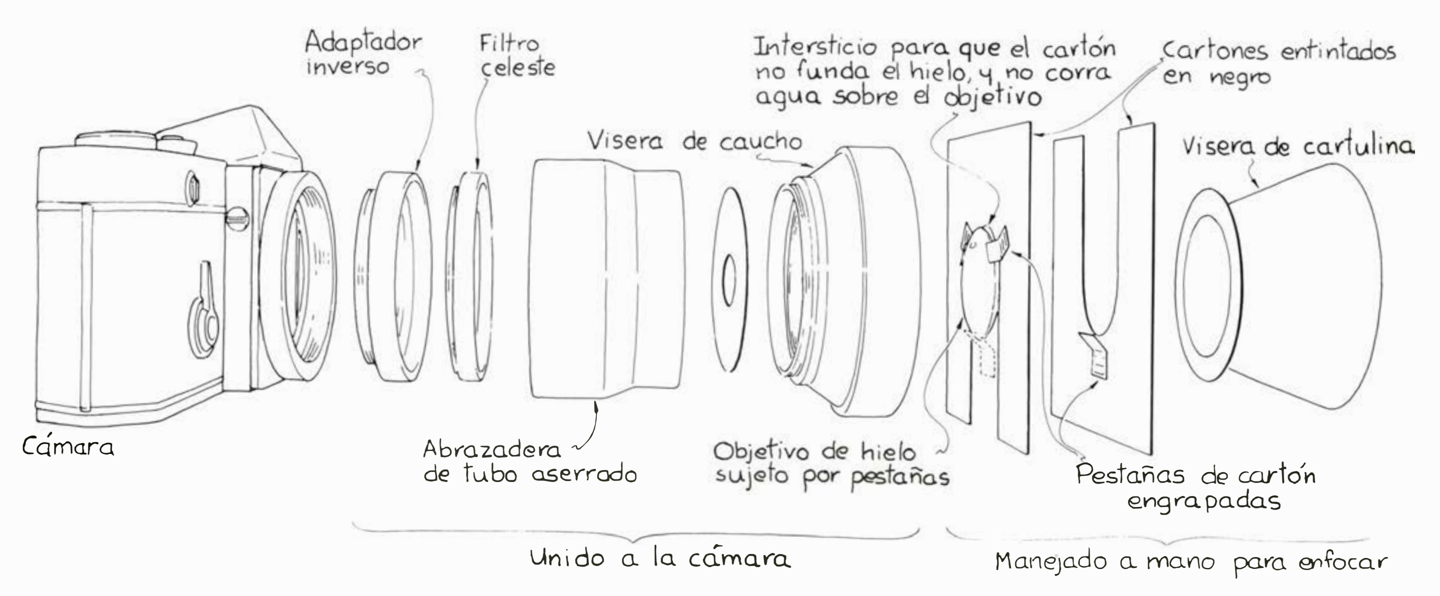
Hay quien mezcla los ingredientes antes de preparar la bebida, pero yo no acostumbro a hacerlo. Tras colocar el ibrik sobre el hornillo de mi cocina eléctrica pongo el mando de calor al máximo. Entonces, la mezcla no tarda en comenzar a susurrar y chasquear. Pasados algunos minutos, el ruido cesa de repente y se inicia la formación de espuma. Esta crece con gran rapidez hacia la boca del ibrik. Dicha transición es tan rápida que a veces no alcanzo a retirar el recipiente del fuego antes de que rebose la espuma. Cuando vierto el café lentamente en la tacita, parte de la espuma persiste, pero la mayoría desaparece.

Con el propósito de observar el interior de la mezcla mientras se cocía, sustituí el ibrik por un vaso de precipitación hecho de Pyrex con una capacidad de 180 centímetros cúbicos. Volví a introducir las cantidades de agua y café correspondientes a una tacita. El café molido flotaba y se humedecía muy lentamente. Parece ser que el polvo se mantiene unido merced a una fuerza eléctrica de intensidad suficiente para que forme terrones, entre cuyos granos el agua se infiltra con dificultad. Así, aunque las partículas sean más densas que el agua, por lo que deberían hundirse, el aire atrapado que contienen los terrones basta para mantenerlas a flote.

Luego añadí dos cucharaditas de azúcar por encima del café. En su mayoría, el azúcar se hundió, a veces volcando algún terrón de café. Así, cuando puse el vaso sobre el hornillo, se distinguían dos capas, una de azúcar y otra de café, perfectamente separadas por una capa de agua bastante transparente.

Sucede que esas tres capas aparecen aun cuando agite la mezcla antes de calentarla. En agua fría, el azúcar se disuelve lentamente y, por ello, se posa en el fondo tras agitar porque es más denso que el agua. En lo que respecta al café, sus partículas, incluso con la agitación, siguen conservando aire adsorbido suficiente para que, en su mayoría, floten hacia la superficie. La única diferencia que supone la agitación es que algunas de las partículas acaban en la capa de agua.

Al poco de empezar a calentar, aparecieron burbujas de aire en distintos puntos del fondo de la capa de azúcar que oscilaban rápidamente. El aire de esas burbujas procedía del adsorbido en los granos de azúcar y del disuelto en el agua que rodeaba a los mismos. Cuando en un lugar determinado se ini-



3. Dispositivo del que se vale Wheeler para montar un objetivo de hielo en una cámara

cia la germinación de una burbuja afluye a ese punto más aire procedente del agua y del azúcar circundantes. Las burbujas oscilan porque se alargan hacia arriba y entran en contacto con líquido más frío (agua azucarada concentrada en este caso) con lo que se contraen repentinamente; pero, casi al instante, vuelven a calentarse y a dilatarse. Esta rápida y desordenada oscilación de las burbujas de aire genera los chasquidos y murmullos que caracterizan la primera etapa del calentamiento de un fluido.

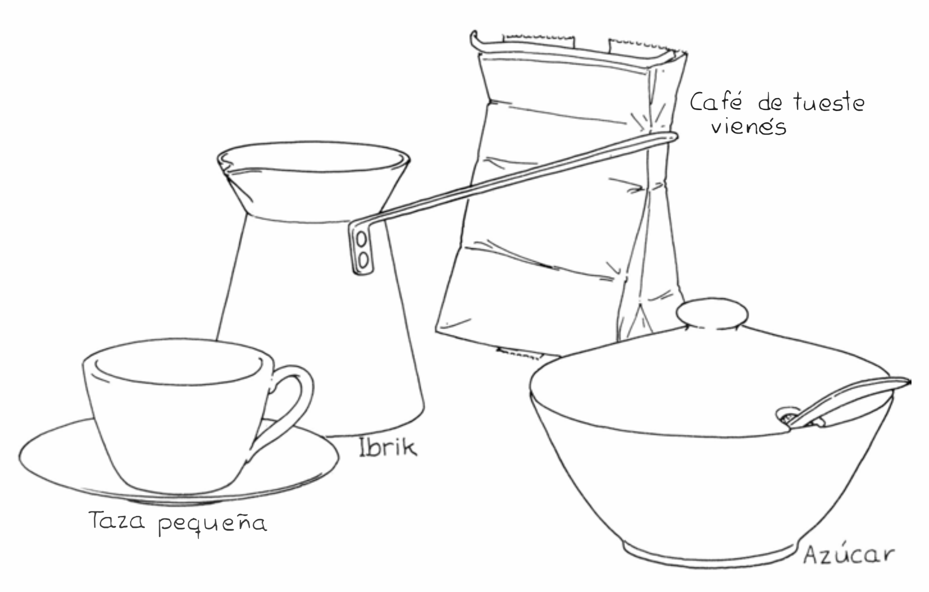
Las burbujas de aire y vapor acaban liberándose y corren hacia arriba atravesando la capa de azúcar. Cuando traspasan ésta y penetran en el agua clara, detienen su movimiento. Esta agua está más fría que la capa de azúcar y hace que las burbujas revienten.

Iluminando el vaso con una linterna me es posible localizar los pequeños chorros de agua azucarada que resultan lanzados hacia arriba al reventar cada burbuja. Esos minúsculos chorros son visibles porque poseen un índice de refracción distinto del característico del agua y se aprecia cómo se retuercen y giran antes de hundirse otra vez en la capa de azúcar. En ocasiones, la estela de un chorrito deja a su paso gotitas de azúcar diminutas (o de agua azucarada).

La capa de azúcar no tarda en tornarse turbulenta. Si bien, hasta ese instante, las burbujas transportaban calor hacia la capa de agua superior, a partir de ese momento parecen predominar las corrientes convectivas, algunas de las cuales se dejan ver merced a su distinto índice de refracción. Al proseguir el calentamiento la capa de azúcar en turbulencia asciende e invade el estrato acuoso. Entonces, parte de la turbulencia o alguna burbuja alcanza alguna que otra vez la base del café; por lo demás, a éste no parece afectarle lo que acontece debajo.

A medida que la capa de azúcar turbulenta se aproxima al café, las partículas de éste comienzan a relucir gracias a diminutas burbujas de aire, ninguna mayor que un milímetro y, en su mayoría, mucho más pequeñas. Tan pronto la capa azucarada alcanza el café, sus partículas comienzan a agitarse y liberan aquellas burbujitas en gran cantidad. Estas forman la espuma, la cual crece a tal velocidad que, en pocos segundos, su altura triplica la de la mezcla inicial.

En cuanto se generaliza la agitación de las partículas, cesan el murmullo y los chasquidos. Este ruido tenía su origen, efectivamente, en la formación de burbujas grandes en el fondo del recipiente; pero al llegar allí los granos ofrecen un sinnúmero de puntos donde pueden germinar burbujas pequeñitas, y desaparecen, por tanto, las ruidosas burbujas grandes.



4. Elementos necesarios para preparar un café turco

Para asir el vaso caliente me valgo de unas tenazas largas. Sosteniéndolo por encima del hornillo, regulo la rapidez a la que se produce la espuma: al alejar el vaso del calor la espuma se genera con mayor lentitud. Las burbujas empiezan pequeñas, pero conforme las nuevas empujan a las anteriores vaso arriba, se fusionan y originan otras mayores. Estas resultan demasiado frágiles para sobrevivir cuando se sirve la bebida. La espuma que a algunos les agrada en el café turco se compone de las burbujas diminutas que estaban adheridas a las partículas de café al iniciarse la formación de espuma. Cuando se sirve el café en su taza, parte de las burbujitas se desprenden de las partículas y forman una especie de masa flotante que puede conservar su configuración durante un buen rato.

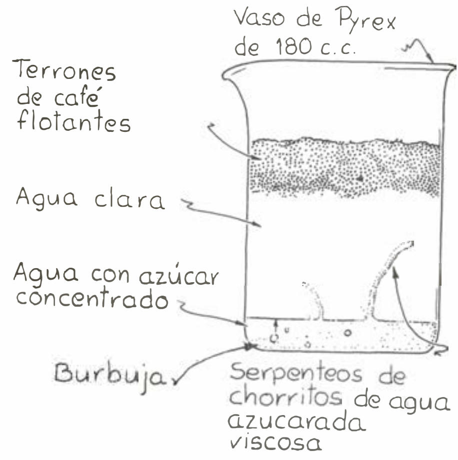
No siempre consigo tal masa espumosa al hacer un café. Pero si acierto, persiste horas enteras mientras no se agite la taza. Una causa de la estabilidad de las burbujas reside en el hecho de que el azúcar acrecienta la viscosidad del fluido. Cuando se forma una masa espumosa, la gravedad hace que el fluido situado entre las paredes de las bolsas de gas se escurra hacia abajo. Así, las paredes se adelgazan hasta que las burbujas revientan o se fusionan. Pero cuanto más viscoso es un fluido tanto más lentamente escurre. El líquido contenido entre las paredes de las burbujas de los cafés que yo preparo tarda mucho en escurrir.

Para ver qué pasaba al cambiar las condiciones puse a calentar un café sin azúcar. Entonces, las partículas volvieron a flotar en terrones sobre el agua y, al poner el vaso en el hornillo, el agua arrancó a hervir rápidamente haciendo mucho ruido. De tal modo que se formaron burbujas de aire oscilantes y, seguidamente, de vapor de agua que escapaban hacia arriba para chocar con la base de la capa de café.

Estas últimas no tardaron en formar corrientes que rompieron los terrones de café, cuyas partículas se dispersaron por todo el volumen de fluido. Los terrones perdieron así parte del aire que contenían y, por ello, capacidad para flotar; pero todavía conservaban aire adsorbido. Entonces, si se hundían era a causa de las fuertes corrientes convectivas que se habían organizado en el vaso. Se formaron burbujas diminutas y comenzó a generarse la espuma. Todo esto sucedió más deprisa que cuando la mezcla contenía azúcar.

Probé con algunas otras combinaciones. En un caso calenté sólo azúcar (dos cucharaditas) y agua. Como antes, el azúcar se hundió en el fondo y se calentó primero. Volvieron a formarse burbujas viscosas y se produjo una convección turbulenta. Cuando la mezcla se encontraba en plena ebullición turbulenta, en la superficie brotaron burbujas, pero aquí no hubo formación rápida de espuma. Las burbujas eran relativamente grandes (de varios milímetros de diámetro) y no duraban mucho.

Pasados algunos minutos, el fluido se tornó transparente, lo que indicaba que el azúcar se había disuelto por completo y mezclado uniformemente. Entonces, dejé que la mezcla se enfriara y observé que todo el azúcar permanecía disuelto, incluso a la temperatura ambiente. Es, pues, evidente que la cantidad de azúcar que normalmente se aña-



5. Infusión de café en un vaso de precipitación

de a la bebida no rebasa el punto de saturación del azúcar en agua caliente.

Después herví en el vaso la cantidad de agua acostumbrada, pero sin azúcar. Cuando se inició la ebullición turbulenta añadí una cucharadita de café. Casi al momento, sus partículas se arremolinaron por todo el volumen de agua y sobre ellas se generaron burbujas minúsculas. La espuma se formó a tal velocidad que parte del líquido se derramó sobre el hornillo antes que alcanzara a retirar el vaso del calor.

En otra combinación reemplacé el azúcar por una cantidad igual de miel. Lo primero que introduje en el vaso fue ésta, seguida del agua y, por último, el café. Cuando la miel empezó a calentarse, de ella surgían minúsculos géiseres que se proyectaban hacia el estrato acuoso. Sólo me era posible verlos iluminando el vaso con una linterna que patentizase la diferencia entre los índices de refracción de la miel y del agua. Cuando algún géiser se desgajaba de la capa principal de miel, volvía a caer en ella: la miel es más densa que el agua. Muchas veces el chorro se enroscaba o soltaba gotas de miel. De éstas, algunas se quedaban donde estaban, por la razón obvia de que se había diluido lo suficiente para adquirir una densidad más o menos igual que la del agua.

En ese punto de mis investigaciones, disponía ya de datos para responder a algunas preguntas que me había formulado sobre la elaboración del café turco. El fundamento de la preparación reside en cocer rápidamente las partículas de café. Si la cocción es excesiva, la bebida adquiere un sabor amargo o ácido; como un café normal que se haya hecho cocer lentamente durante mucho tiempo. Para que se calienten y cuezan rápidamente, se muelen finamente los granos para café turco.

La espuma constituye casi un efecto secundario. Las partículas diminutas ofrecen mucha superficie donde adsorber aire. Las minúsculas bolsas de aire son puntos de germinación de burbujas. Entonces, cuando el café y el agua circundante se calientan, el aire disuelto en ésta sale de solución en los puntos de nucleación y forma burbujas de un tamaño apenas suficiente para que sean visibles.

Cuando la capa de líquido caliente llega a las partículas, el vapor de agua dilata las burbujas; éstas se desprenden y separan de los puntos de nucleación para formar la espuma de burbujas grandes que remonta el ibrik. Un gran número de los puntos de nucleación existentes en las partículas permanecen activos, pues retienen parte del agua o vapor cuando una burbuja se separa de ellas. Y, además, la agitación los transporta de un punto a otro del agua contigua al fondo del recipiente y prosiguen así generando burbujas en tanto el fluido circundante esté suficientemente caliente para proporcionar vapor de agua.

El azúcar se añade antes de la cocción. Hay varias razones para obrar así. El calentamiento asegura que el azúcar se halle totalmente disuelto en el momento de beber el café. Si se agregara después de servir el café en la taza, habría que agitar el líquido para disolverlo, lo cual supondría que las partículas permanecerían en suspensión durante mucho más tiempo. Y, así, el café se enfriaría antes de que se hubieran posado lo suficiente para que la bebida no resultase pulverulenta.

Ocurre, además, que el azúcar facilita la buena cocción. En efecto, la capa viscosa de azúcar que se forma en el fondo del ibrik retiene el calor mejor que el agua, de tal modo que, si dicha capa contuviera sólo agua, las burbujas de vapor transportarían calor hacia el café en una etapa más temprana del proceso. Pero, gracias al azúcar, eso se retarda y el café tarda más en empezar a calentarse; y ello merced a la capa de agua azucarada en turbulencia que acaba llegando a él. El período de calentamiento y cocción del café es repentino y breve; se inicia cuando le llega el líquido caliente y finaliza en menos de un minuto, cuando la espuma amenaza con rebosar del recipiente.

Por último, el azúcar es beneficioso para el proceso porque favorece la estabilización de la espuma de burbujitas. Sin su presencia, sería menos probable que la espuma persistiera tras servir la bebida en la taza.

Aunque el proceso de infusión sea rápido, no debe concedérsele un tiempo demasiado corto. Cuando utilizo el ibrik para hacer a la vez dos tazas de café, suele quedarme pulverulento. Por más que espere, las partículas grandes no llegan a posarse en el fondo. Ello se debe a que todavía encierran aire adsorbido en cantidad suficiente para permanecer a flote; cabe, incluso, que sigan apelmazadas. Como el ibrik contiene demasiado líquido, me es necesario retirarlo prematuramente del hornillo para que la espuma no rebose. En consecuencia, el tiempo que transcurre desde que la capa azucarada, caliente y en turbulencia, llega al café y el momento en que éste se sirve en la taza resulta demasiado corto para que se elimine todo el aire adsorbido. Mi experiencia práctica no me permite preparar más de una taza de café cada vez. Cuando es necesario hacer una cantidad mayor en una sola preparación, debo sostener el ibrik separado de la fuente de calor para controlar la espuma y mantener la cocción el tiempo suficiente.

Lo que viene ahora atañe a lo que acontece en la taza cuando las partículas de café se hunden hacia el fondo. Las partículas se hunden todas a una velocidad aproximadamente constante. Dada la importancia, tanto científica como tecnológica, que reviste la velocidad de sedimentación de partículas en suspensión, se han realizado muchas investigaciones acerca del tiempo que tardan en posarse partículas en el seno de agua.

Si las partículas son menores que 0,06 milímetros aproximadamente, caen obedeciendo a la ley de Stokes, llamada así en honor del físico y matemático británico G. G. Stokes. La velocidad de una partícula pequeña depende de dos factores: su peso efectivo y la resistencia que opone al agua. El peso efectivo es la diferencia entre el peso verdadero de la partícula y el empuje de Arquímedes que le comunica el líquido que la rodea, el cual suele expresarse en función de las densidades de la partícula y del fluido. Como las partículas de café son más densas que el agua (y más densas que la mezcla de azúcar y agua), el empuje que reciben es demasiado débil para impedir que caigan. O sea, el peso de la partícula es un vector apuntado hacia abajo más corto que el peso verdadero.

Oponiéndose a esta fuerza, el fluido crea una resistencia al avance que se materializa en una fuerza dirigida hacia arriba. Según la ley de Stokes, la resistencia al avance depende del diámetro de la partícula, de la velocidad de ésta y de la viscosidad del fluido. Cuando la

partícula inicia su descenso, su velocidad es reducida y, por tanto, también lo es la resistencia. Pero la partícula se acelera y en seguida adquiere una velocidad suficiente para que la resistencia al avance iguale al peso efectivo. A partir de entonces, la partícula ya no se acelera, pero sigue cayendo a velocidad constante. Esta velocidad final depende del tamaño de la partícula. Las más pequeñas tienen unas velocidades finales menores y tardan más en posarse.

El valor de la velocidad final, que en física se conoce como velocidad límite, depende asimismo de la viscosidad del fluido. Como el azúcar la incrementa, todas las partículas tienen unas velocidades límite inferiores a las que les corresponderían en ausencia del mismo.

Infortunadamente para los degustadores del café turco, la ley de Stokes es válida sólo para partículas esféricas. Las partículas de café suficientemente pequeñas para que les sea aplicable la ley de Stokes son de forma irregular y, por tanto, encuentran una resistencia difícilmente calculable. Aún así, se mantiene la generalidad de la influencia del tamaño de la partícula: caen más lentamente cuanto más pequeñas son.

Gran cantidad de las partículas del café turco son excesivamente grandes para caer obedeciendo a la ley de Stokes, la cual establece que la fuerza de resistencia al avance depende de la primera potencia de la velocidad de la partícula respecto al fluido. Para partículas mayores, esa fuerza depende del cuadrado de dicha velocidad. Además, la expresión matemática de la fuerza incluye un coeficiente de resistencia a determinar experimentalmente, el cual tiene en cuenta el efecto de la forma irregular de la partícula.

Otro factor es que si en un volumen de fluido determinado caen un gran número de partículas, la situación varía de dos maneras. Por una parte aumenta la densidad efectiva del fluido y, por consiguiente, disminuye el peso efectivo de cada partícula y su velocidad de sedimentación. Por otra parte, las partículas interactúan entre ellas al pasar unas junto a otras en su camino hacia el fondo. Pese a estas complicaciones adicionales, conserva su validez la regla anterior: las partículas caen más lentamente cuanto más pequeñas son.

Para verificar la regla introduje una pequeña cantidad de café turco recién hecho en una jarra alta y transparente que contenía agua. Esta la había dejado sin mover durante una hora con el propósito de que estuviera muy encalmada. Utilicé café recién hecho, pues necesitaba granos que hubieran perdido la mayor parte del aire adsorbido. Unos granos no cocidos hubieran tendido a flotar.

Valiéndome de una linterna pude seguir el descenso de las partículas. Las grandes se hundía rápidamente hacia el fondo a una velocidad del orden de un centímetro por segundo. Las pequeñas eran mucho más lentas. Y algunas de las más pequeñas parecían invertir 10 minutos o más en llegar al fondo.

La velocidad de sedimentación de las partículas en una taza de café difiere algo de la observada en la comprobación que acabo de describirles. El número de partículas por unidad de volumen es suficientemente alto para que interactúen mientras caen. Y todavía más importante, el fluido está lo bastante caliente para crear células de circulación. Lo que ocurre es que el líquido caliente contiguo al fondo se eleva hacia la parte superior, donde se enfría por evaporación y luego desciende. Esta circulación forma las llamadas células de Bénard. En ciertos fluidos y bajo determinadas condiciones de calentamiento, las células crean una configuración geométrica estable en la superficie.

En los cafés por mí preparados esa configuración es caótica y constantemente variable. Se observa mejor cuando el café carece de espuma, pues ésta no sólo cubre la superficie sino que altera la velocidad de evaporación. Instalando una luz que se refleje en la superficie, se descubren minúsculas gotitas suspendidas inmediatamente por encima de la superficie de la zona de fluido caliente ascendente. Estas gotitas son el producto de la condensación del vapor de agua cuando se enfría el aire inmediatamente contiguo a la superficie. En la zona de fluido descedente no se ven gotitas. Como éstas reflejan la luz, las zonas de fluido ascendente parecen levemente blancas, mientras que las de fluido descendente presentan el color oscuro del café. En el caso de una taza de café carente de partículas, las células de Bénard no son más que una curiosidad; pero en el caso del café turco suponen un retraso en la sedimentación de las partículas.

La conclusión es que, al tomar café turco, lo recomendable es buscar una solución de compromiso, esperando lo suficiente para que se posen en el fondo las partículas de café mayores y para que decrezca la circulación de Bénard, pero no tanto que la bebida se enfríe hasta un punto no agradable. La espuma facilita ese compromiso, aislando el café mediante el vapor atrapado en las burbujas y haciendo así que disminuya la velocidad de enfriamiento.

Libros

Atlas de histología comparada, multiplicación génica, física subatómica y jardines de Kew

J. J. Vázquez, Tomás Santos, J. P. Acordagoicoechea y F. de Diego

Introducción a la Histología Animal Comparada. Atlas-Libro de la Estructura Microscópica de los Animales, por Ricardo Paniagua Gómez-Alvarez y Manuel Nistal Martín de Serrano, con la colaboración de M.ª Asunción Abaurrea Equisoain. Editorial Labor, S. A., 438 páginas, 669 figuras; Barcelona, 1983. Dos etapas jalonan el avance en el conocimiento de la histología animal comparada a lo largo de nuestro siglo. En la década de los 20 se registra la primera aportación sistemática, realizada por zoólogos, que se recopila en obras publicadas en torno a los cincuenta, como la *Histología Comparada* de Andrew, que han mantenido su vigencia hasta tiempos recientes. En los últimos 20 años se ha vuelto a vivir un renacer de la histología animal comparada, con numerosas publicaciones en revistas especializadas, que han supuesto un considerable incremento de información, que exige una síntesis que establezca los patrones básicos de diseños y las desviaciones producidas por la adaptación a las diferentes circunstancias ecológicas y evolutivas.

A lo largo del último decenio han surgido varias obras de carácter comparado, que, si no se limitan a los vertebrados, ponen el acento en ellos, especialmente en mamíferos. Se echaba en falta una recopilación de los numerosos conocimientos adquiridos en el grupo de los invertebrados.

En el presente manual, escrito por Ricardo Paniagua y Manuel Nistal, con la colaboración de Asunción Abaurrea, se contempla todo el reino animal, con especial atención a los invertebrados. El espacio dedicado a los mamíferos es relativamente reducido por considerar los autores, con buen criterio, que la histología de estos últimos se halla suficientemente representada en los manuales de histología existentes. Los autores tienen amplia experiencia histológica. Paniagua y Nistal son profesores agregados de la materia y autores de numerosas publicaciones en revistas de prestigio mundial. Abaurrea es profesor adjunto y domina el arte histológico, poseyendo una alta calidad técnica.

El libro está dividido en 13 capítulos. El primero trata muy someramente de las células y los tejidos, a modo de recordatorio. Al hablar de los tejidos, se describen, en cambio, algunos aspectos comparados interesantes. Los 10 capítulos que siguen se dedican a los diferentes sistemas orgánicos: nervioso, circulatorio, linfoide, respiratorio, digestivo, excretor, endocrino, reproductor masculino, reproductor femenino y tegumentario. Los dos últimos capítulos tratan de los neurorreceptores.

Cada capítulo suele comenzar por unas consideraciones sobre la estructura general y las implicaciones fisiológicas. Se aborda, luego, el fenómeno histológico según un orden filogenético ascendente, yendo generalmente de menor a mayor grado de complejidad. Cuando el sistema está presente, no falta nunca la descripción en Anélidos –preferentemente en Oligoquetos–, Artrópodos y Moluscos –Gasterópodos en particular. Al final del capítulo aparece una relación de referencias bibliográficas en las que el lector puede ampliar algunos aspectos concretos, especialmente ultraestructurales, ya que ha sido intención de los autores limitarse al ámbito de la microscopía óptica, para conseguir un libro manejable.

El texto, bien redactado, es claro y conciso, lo que permite su cabal comprensión. La iconografía es abundante, como exige un manual de histología y más aún de histología comparada, dada la diversidad de patrones existentes en invertebrados. Se multiplican las fotomicrofotografías, los dibujos y los esquemas. No todas las micrografías alcanzan la misma calidad, si bien la mayoría son aceptables, y algunas, de cortes semifinos, excepcionales. Los esquemas y dibujos, seleccionados con acierto, contribuyen poderosamente a la comprensión de los cuadros histológicos. En alguna ocasión quizá se echa en falta algún esquema más, o alguna imagen panorámica que nos ofrezca una representación de conjunto.

El formato del libro es muy grato y manejable; la impresión, de buena calidad. La ordenación de las figuras preferentemente en las páginas impares y el texto en las pares constituye un gran acierto editorial, ya que facilita notablemente la lectura del texto y la consulta simultánea de las figuras.

Algunas interpretaciones discutibles y ocasionales errores de poca importancia, casi siempre en pies de figuras, debidos probablemente al marcado, constituyen los únicos aspectos criticables, pero no restan valor a este interesante y meritorio manual, que será de gran utilidad para estudiantes y profesores de las facultades de ciencias biológicas. De valor asimismo para cuantos se interesan por la histología animal y humana. (J. J. V.)

Gene Amplification, dirigido por R. Schimke. Cold Spring Harbor Laboratory; C.S.H. New York, 1981. Hasta la década de los setenta, la posible existencia de procesos de multiplicación (o amplificación) génica se reducía, dentro del cuerpo de doctrina de la biología, a algunas observaciones esporádicas en el dominio de la genética formal (frecuentemente marginada cuando no menospreciada en la actualidad) que se remontan a los clásicos experimentos de Sturtevant en *Drosophila*. Este investigador recabó, ya en 1925, una curiosa propiedad del locus "bar", en el cromosoma X, cual era que dicha mutación revertía con una frecuencia anormalmente elevada y concluyó que la reversión se debía a "un cambio cuantitativo en el gen bar".

Con la demostración por Brown y Dawid en 1969 de la multiplicación de genes para el ARN ribosómico en el sapo africano (*Xenopus laevis*), en el marco de sus estudios sobre biología del desarrollo, la posibilidad de la existencia, más que esporádica, de procesos de multiplicación o amplificación génica en el mundo vivo comienza a recibir una atención creciente. Finalmente, el advenimiento de las técnicas de manipulación "in vitro" del ADN determina en los últimos cinco años una proliferación explosiva de trabajos procedentes de distintos ámbitos de estudio, particularmente de la citogenética y biología molecular, que demuestran la existencia de multiplicación génica en varios sistemas de células eucariotas. Con el ánimo de contrastar y unificar estas últimas y dispersas observaciones, procedentes de campos frecuentemente muy variados, se celebró una conferencia sobre multiplicación génica en Cold Spring Harbor, en octubre de 1981, fruto de la cual es este li-

bro que aglutina, por primera vez, en un cuerpo coherente, los datos más recientes en torno a este sorprendente fenómeno.

El libro es, pues, una recopilación monográfica de las aportaciones de los distintos conferenciantes, que lleva el sello de esta prestigiosa editorial, destacada promotora de revisiones, siempre recomendables, de temas de actualidad, principalmente de genética y biología molecular. El principal interés del libro reside, como comenta su editor (y destacado investigador de este campo) Robert Schimke, en que brinda, recopiladas, todo un cúmulo de pruebas que permiten concluir que la multiplicación génica es un mecanismo evolutivamente consolidado de *modulación de la expresión génica*, aportación ésta a todas luces innovadora si se considera que probablemente gran número de estudiantes de biología todavía escuchan en las aulas que la regulación de la expresión génica opera modulando la velocidad de transcripción o traducción, o bien controlando la estabilidad o el procesado de los productos de transcripción (ARN) o traducción (proteínas), en la concepción hasta hace muy poco vigente del código genético como un texto rígido, cuya expresión no exigía ninguna modificación del mismo.

El contenido de los artículos que se recogen en el libro incide, pues, en un aspecto que obliga a ampliar el esquema que resume los puntos de control potencial de la expresión del mensaje genético, incluyendo la propia modulación del número de copias de un gen como mecanismo que garantiza el aporte, en las cantidades requeridas, de un producto génico cuya demanda aumenta en unas condiciones determinadas o en un momento del desarrollo, en aquellos genes donde la evolución no ha proporcionado un mecanismo eficaz para modular su lectura según las necesidades del momento.

Sin menosprecio de la utilidad indiscutible del libro, su organización adolece de cierta falta de sistematización que obliga a realizar un esfuerzo adicional para obtener una visión ordenada y realmente aprovechable de su contenido, aspecto éste que hubiera sido subsanable con una mayor atención al mismo. Sistematizando, pues, su contenido, el lector puede obtener valiosa información en torno a tres puntos básicos de la multiplicación génica: situaciones que determinan la multiplicación de genes específicos en un organismo, estructuras moleculares de las secuencias multiplicables y fenomenología de la multiplicación y, finalmente, sus mecanismos moleculares.

En el primer tercio del libro se recogen los trabajos que describen situaciones que conducen a la multiplicación génica y que se podrían resumir en las siguientes: respuestas adaptativas a una gran presión selectiva, tales como la presencia de fármacos inhibidores del crecimiento que determinan la multiplicación de los genes que rigen el producto afectado por el inhibidor. En esta línea, el mayor número de aportaciones, de calidad desigual, se refieren al caso tal vez más conocido en la actualidad: multiplicación de los genes para la dihidrofolatorreductasa, que determina la selección de células resistentes a metotrexato, un fármaco ampliamente utilizado en clínica como antitumoral que por esta precisa razón, la de inducir multiplicación génica, suele conducir al desarrollo de resistencias que limitan su utilidad clínica. De estos trabajos, tal vez los más completos y de mayor interés sean el de Bostock (pág. 15) y el de Mary C. Weiss con el propio Robert Schimke (pág. 29); describen las características citogenéticas y moleculares que manifiestan los distintos variantes fenotípicos donde ha ocurrido la multiplicación de dichos genes.

El segundo de los casos, documentado mediante cuatro trabajos, se refiere a la multiplicación de genes, controlada por el propio organismo, durante una etapa del desarrollo, como mecanismo que garantiza el aporte de un producto (codificado por el gen) en las cantidades requeridas ante una gran demanda específica en ese período particular del desarrollo. En dos de dichos trabajos, el de A. C. Spradling (pág. 121) y el del brillante grupo de Kafatos (pág. 129), se presentan los últimos resultados en torno a la multiplicación y expresión de genes para las proteínas del corion en *Drosophila*, procesos cuidadosamente controlados en una etapa muy restringida del desarrollo embrionario. Igualmente ilustrativo y novedoso resulta el trabajo de Zimmer (pág. 137) sobre la multiplicación de genes de la actina durante la miogénesis en el pollo. La tercera de las situaciones donde se ha descrito multiplicación génica tiene que ver con procesos cancerígenos, representada por varios trabajos que inciden en la presumible importancia de la multiplicación génica en el desarrollo de algunos tumores. De ellos son particularmente meritorios el de Hubbell (pág. 193), el único que, mediante clonaje molecular, analiza con relativo detalle la naturaleza de las secuencias, y el de Barsonm y Varshavsky (pág. 239) que demuestran cómo los agentes cancerígenos inducen un aumento considerable en la frecuencia de multiplicación génica. En el trabajo de Richard Axel (pág. 251) se demuestra la utilidad de la técnica de transfección de células en estudios sobre multiplicación y se pone de manifiesto el papel de la misma como mutación reversora.

El segundo bloque de trabajos, de contenido heterogéneo y a menudo solapados con alguno del primer tercio, incide en el análisis de la estructura molecular de las secuencias repetidas. A pesar de ser reciente la exploración del problema, estos trabajos ilustran el potencial de las técnicas de análisis molecular, que han permitido, en cuestión de meses, obtener una imagen bastante completa de la estructura fina de las distintas secuencias multiplicadas, hasta el punto de que, de todos estos trabajos, se puede inferir un modelo general que resume la fenomenología del proceso de multiplicación, particularmente documentado en el trabajo de Levan (pág. 911), que comienza con la replicación diferencial de un segmento de ADN, de longitud variable en cada caso, que se manifiesta por el alargamiento de una región del cromosoma (regiones HSR). Esta región repetida puede escindirse para originar entidades autónomas autorreplicativas conocidas por minicromosomas (DMs), que pueden de nuevo integrarse, mediante recombinación homóloga, en la región HSR del cromosoma o incluso en un cromosoma distinto, conduciendo, pues, a reorganizaciones génicas.

De este bloque de trabajos destacan, por lo avanzado de sus análisis, los de Robert Schimke y colaboradores (pág. 149) sobre la variabilidad de las estructuras repetibles del gen de la dihidrofolatorreductasa en células de ratón y el notable estudio del grupo de Wahl (pág. 167) sobre las estructuras que contienen las secuencias del complejo de la carbamoilfosfatosintetasa-aspartatotranscarbamilasa y dihidroorotasa (complejo CAD).

A continuación sigue un grupo de trabajos que, aunque se engloban bajo el epígrafe "Mecanismos de multiplicación génica" en realidad es, de nuevo, de contenido heterogéneo, ya que un auténtico análisis sobre los mecanismos moleculares que operan en el proceso se realiza sólo en el trabajo de Bullock (pág. 215), quien pormenoriza los acontecimientos moleculares en la escisión y multiplicación del virus SV40 integrado en células de mamífero, y en uno anterior de Hamlin (pág. 155), donde se estudia el proceso (y se propone un modelo) sobre el mecanismo de multiplicación de los genes de la dihidrofolatorreductasa. Especial mención merecen en este contexto los tra-

bajos de Philip Sharp (pág. 245) y Richard Axel (pág. 251), donde se resalta, de nuevo, la utilidad de transfección del ADN, en la dilucidación de los mecanismos de la multiplicación.

El libro termina con un corto apartado sobre alteraciones cromosómicas, orientado a establecer su papel en procesos tumorales, y con una acertada revisión y evaluación final por Robert Schimke del estado actual de este fascinante campo, así como de las incógnitas aún pendientes que obligatoriamente marcarán las líneas de investigación a corto plazo en este nuevo filón de la genética molecular.

Disculpando la superficialidad de algunos trabajos y los defectos formales, particularmente la arbitraria organización de los distintos apartados, el libro, en resumen, resulta de indudable interés para estudiosos no sólo de la genética sino de la biología fundamental, ya que ofrece una recopilación del creciente número de observaciones en torno a este novedoso y atractivo descubrimiento de la biología, cuya principal enseñanza tal vez sea la de contribuir a modificar la concepción tradicional del genoma como un archivo estático religiosamente protegido, modificable por error o accidente, por la más verosimil y actual donde se perfila el extraordinario dinamismo de las moléculas de ADN genómico, sujeto a variadas reorganizaciones, enzimáticamente controladas, dentro de la particular estrategia de cada especie en la expresión de su programa de desarrollo. (T. S.)

PARTÍCULAS, CAMPOS Y SIMETRÍAS, por Manuel García Doncel. Publicaciones de la Universidad Autónoma de Barcelona (UAB), Bellaterra, Barcelona, 1982. Este trabajo, segundo de la serie de historia y epistemología de la física (HEF) publicado por la UAB (el primero *Relatividad Especial, Relatividad General (1905-1923)*, por José M. Sánchez Ron, está agotado y es inminente su publicación en Alianza Editorial) presenta la historia de la física de altas energías desde su nacimiento (en la década de los 30) hasta los 60. Esta exposición histórica, simplificada por motivos didácticos, está dividida ("artificialmente", dice el autor, p. 12; si bien la mutua implicación queda suficientemente señalada y recogida en el trabajo) en tres partes o capítulos, que corresponden a las tres palabras del título: partículas, campos y simetrías.

En la primera (pp. 17-73) se estudia la historia del descubrimiento de las nuevas partículas, que van cambiando el mundo de los años 30, constituido por sólo tres partículas (electrón, protón y fotón), hasta la complejidad de nuestro mundo actual, que engloba a centenares de partículas, entre las que se incluyen las partículas subatómicas, las antipartículas y las llamadas resonancias, ya que sus propiedades físicas permiten identificarlas como partículas. Es un capítulo más fenomenológico y menos teórico que los dos restantes.

En la segunda parte (pp. 75-144) se aborda la historia de la introducción de los diversos tipos de campos cuánticos, comenzando por el campo de fotones, introducido por Dirac en 1927, el de electrones, introducido en 1928 por Jordan y Wigner, y los campos que corresponden a las partículas descritas en el capítulo anterior. Sigue una exposición de la electrodinámica cuántica relativista de los años 30, que fue reformulada de forma "elegante" tras los experimentos con microondas realizados durante la segunda guerra mundial. En las secciones 5 y 6 de este capítulo se trata la extensión de la teoría cuántica de campos a otras interacciones distintas de las electromagnéticas: a las interacciones nucleares débiles por parte de Fermi, en 1933, y a las fuertes, realizada por Yukawa en 1935.

Se presenta en la tercera parte (pp. 145-211) un recorrido histórico de la introducción de los diversos tipos de simetría que interesan en este mundo subatómico. Punto de partida obligado es el modelo de simetría introducido por la relatividad especial. (Einstein nos enseñó a ver la simetría como invariancia; en el caso de la relatividad especial, como invariancia bajo traslaciones y rotaciones espacio-temporales.) Ello exige remontarse, aunque sea sucintamente, a un período anterior al estudiado en el trabajo. A continuación se expone el desarrollo histórico de la introducción de otras simetrías más "sutiles", que se relacionan con el espín (momento angular interno) y la estadística. El apartado siguiente se dedica a las simetrías "discretas": la especular (inversión espacial), la temporal (reversión temporal o cambio del sentido de movimiento) y la conjugación de cargas (conjugación partícula-antipartícula). El autor hace ver cómo se fueron descubriendo estas simetrías (hacia los años 40) y cómo (a partir de los 60) se ha ido constatando experimentalmente que en la naturaleza no se satisfacen. Por último, trata de la introducción de las simetrías internas que, atendiendo a sus propiedades de tipo carga, parecen apuntar a una constitución interna de las partículas llamadas "elementales".

En este trabajo M. G. Doncel detiene su estudio en la década de los 60, pues "hemos de temer nos falte todavía la perspectiva histórica necesaria para valorar tales avances [los de los últimos quince años], y distinguir lo que ha sido razonablemente aceptado a prueba, de lo que ha sido razonablemente contrastado en la experiencia y merece pasar a la didáctica y a la historia" (p. 8). Con todo, el mismo autor ha publicado, con carácter algo más divulgador, una actualización del tema hasta 1978 (M. García Doncel: "La física de altas energías", en *Historia de la ciencia 4, Edad Contemporánea*, dirigido por F. Cid, Planeta, Barcelona, 1982; 139-163).

Concluye el trabajo con una valiosa bibliografía (pp. 213-276) agrupada en tres bloques: fuentes primarias (donde se recogen las referencias bibliográficas completas de unos 450 artículos; probablemente la primera vez que aparecen recogidas todas ellas en una sola obra), antologías de fuentes primarias y fuentes históricas. Algunos fragmentos significativos de estos artículos se hallan transcritos en el texto, y en la bibliografía se indica el párrafo en que es citada la correspondiente fuente primaria o histórica.

El trabajo ofrece una excelente y matizada visión de conjunto, que puede ser de mucha utilidad para encuadrar cualquier estudio ulterior sobre temas puntuales. Al mismo tiempo, el autor ha conseguido presentar el desarrollo histórico de la física de altas energías prescindiendo de formalismos matemáticos, lo cual lo hace asequible a un público algo más amplio, y posibilita que el lector interesado, aunque no domine el tecnicismo matemático, conozca y comprenda un campo de investigación puntera actual.

Esperamos que el departamento de publicaciones de la UAB siga con esta interesante serie, pero también sería de desear que se eliminaran algunos yerros que desdicen un tanto de la precisión del trabajo, p. ej.: el maquetaje de las páginas 14-15; la diversidad de criterio en separar los apartados dentro de los capítulos; la poca consistencia de la encuadernación y endeblez del encolado; las discrepancias en la paginación (cfr. pp. 5 y 215), etcétera. (J. P. A.)

KEW: GARDENS FOR SCIENCE AND PLEASURE, editado por F. Nigel Hepper. Stemmer House, Publ. Inc. Owings Mills, Maryland, 1982. De entrada, el libro da la sensación de una cuidadosa y esmerada presentación, con una preciosa fotografía aérea panorámica de una de las partes más representativas de Kew Gardens, la que recoge el Invernadero de las Palmas (Palm House) y sus alrededores.

E. A. Bell, director del Royal Botanic Garden, da cuenta en el prólogo de los principales departamentos, laboratorios y servicios que constituyen estos jardines, así como de su contenido en pliegos de herbario, banco de semillas, museos, plantas de interés económico, volúmenes de la biblioteca (más de 750.000), colección de plantas vivas, etcétera. En general, los datos históricos que se aportan son suficientes y correctos, así como el énfasis en las personas más relevantes en el mantenimiento y expansión de este jardín botánico. Destacan dos por la labor realizada en la primera época: J. Banks (1772-1820) y W. J. Hooker (1841-1865), con un intervalo de 20 años de franca decadencia. Al texto, escrito de forma amena, acompaña una copiosa ilustración, en color y blanco y negro.

La introducción relata la historia de la creación del jardín a partir del que existía en 1772, a la orilla del Támesis, propiedad de la princesa Augusta madre de Jorge III. De las 3,6 hectáreas iniciales el actual ocupa una extensión de 120 hectáreas, visitadas cada año por más de un millón de curiosos y científicos. La colección de especies y variedades de plantas allí presentes rondan las 50.000.

Bajo el epígrafe "Jardines para el placer", se recoge la impresión del visitante un día de primavera, con la evocación del "Bedding", como una moda o estilo de jardín propia del siglo XIX. Este tipo de jardín se produjo a raíz de la revolución industrial, que animó a muchos nuevos ricos a comprar un terreno donde instalar su casa y rodearla de un jardín con la mejor colección posible de plantas valiosas. Ello incitó la búsqueda, por parte de los propietarios, de nuevas variedades mediante intercambio e hibridaciones, consiguiéndose infinidad de nuevos ejemplares que se denominaron "cultivares". Una de las misiones principales de estos "Beddings" era dar un cierto colorido a las plantaciones durante el verano, para después guardar las plantas en invernadero durante el resto del año.

Se nos recuerda luego la creación del Jardín de la Reina en el siglo XVII. Lo consiguió la familia de Jorge III de la Dutch House, construida por Samuel Fortrey en 1631 y más tarde transformada en el Palacio de Kew. El Parterre, Jardín de Plantas del siglo XVII y el Jardín de Sir Henry Price en Wakehurst, con sus extraordinarias y bien cuidadas plantaciones, completan este capítulo.

Recorremos la colección de plantas vivas a través de apartados dedicados a las plantas de rocalla, alpinas, herbáceas, a los invernaderos en Kew y su desarrollo (plantas exóticas, plantas nuevas conseguidas), etcétera. Al presentar los invernaderos, el autor reseña la historia de los principales, su arquitectura y riqueza vegetal. En estos se hacen agrupaciones por regiones geográficas, por familias y por tipos de ecología.

Wakenhurst Place es otro jardín, situado hacia el sur de Kew, a hora y media de distancia, con abundantes lagos, torrentes, zona de cultivos y extraordinaria colección de rododendros. Al explicar la vida silvestre en Kew y Wakehurst Place, conocemos los animales más vistosos que habitan estos jardines. Cuando se aborda la investigación botánica, asistimos a la búsqueda de plantas útiles, entre las que menciona el "árbol del pan" (*Artocarpus altilis*) que, siendo nativo de la zona occidental del Pacífico, se extendió por toda América central, por el interés de sus frutas ricas en carbohidratos. Una planta de importancia clínica es el "árbol de la quina" (*Cinchona officinalis*), fundamental en el tratamiento de la malaria. El "árbol del caucho" (*Hevea brasiliensis*) se destina a usos industriales. Estas y otras muchas plantas de interés para el hombre se cultivan en Kew y desde aquí son distribuidas a otros países para su implantación. Nos enteramos de la organización de expediciones botánicas, su necesidad, quién las dirige, a dónde ir y transporte utilizado. Se habla del papel de Kew en el desarrollo de estudios biológicos y elaboración de floras regionales. Sin pasar por alto la clasificación e identificación de especies, su archivo en los herbarios y conservación de las mismas.

Se está investigando allí, de un modo particular, el género *Crocus*, una de cuyas especies es el azafrán. Sirve de ejemplo para presentarnos un estudio taxonómico a través de las distintas técnicas: cariológicas, citológicas, anatómicas, embriológicas, ultramicroscópicas y palinológicas.

Desde 1969 funciona un banco de semillas que tiene por misión aumentar las posibilidades de conservación durante su almacenaje y facilitar así el intercambio de las mismas con otros jardines botánicos. Se intenta evitar la desaparición de especies en peligro de extinción. En su lucha contra la destrucción de la naturaleza, destaca la búsqueda de nuevas especies capaces de adaptarse a zonas áridas.

Kew: Gardens for Science and Pleasure constituye un trabajo digno de ser tenido en cuenta por todo profesional o aficionado a la botánica, la jardinería o la estética paisajística. (F. D. C.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

UNA ZONA DESNUCLEARIZADA EN EUROPA

COMPREHENSIVE STUDY OF THE QUESTION OF NUCLEAR-WEAPON-FREE ZONES IN ALL ITS ASPECTS. Informe especial del Comité de Desarme de la Conferencia de las Naciones Unidas. Naciones Unidas, 1976.

ARMS CONTROL: THE REGIONAL APPROACH. Ronald G. Purver, Center for International Relations, Queens University, Kingston, 1981.

STUDY ON ALL THE ASPECTS OF REGIONAL DISARMAMENT. Centro de las Naciones Unidas para el Desarme. Informe para la Secretaría General, Naciones Unidas, 1981.

COMMON SECURITY: A BLUEPRINT FOR SURVIVAL. Independent Commission on Disarmament and Security Issues. Simon and Schuster, 1982.

DISPOSITIVOS MICROMECANICOS DE SILICIO

SPECIAL ISSUE ON SOLID-STATE SENSORS, ACTUATORS, AND INTERFACE ELECTRONICS. *IEEE Transactions on Electron Devices*, vol. ED-26, n.° 12, diciembre, 1979.

SILICON AS A MECHANICAL MATERIAL. Kurt E. Petersen en *Proceedings of the IEEE*, vol. 70, n.° 5, págs. 420-457; mayo, 1980.

SPECIAL ISSUE ON SOLID-STATE SENSORS, ACTUATORS, AND INTERFACE ELECTRONICS. *IEEE Transactions on Electron Devices*, vol. ED-29, n.° 1; enero, 1982.

ESTRUCTURA DE QUARKS Y LEPTONES

THE KEY TO THE UNIVERSE: A REPORT ON THE NEW PHYSICS. Nigel Calder. British Broadcasting Corporation, 1967.

TEORÍAS GAUGE DE LAS FUERZAS ENTRE PARTÍCULAS ELEMENTALES. Gerad 't Hooft en *Investigación y Ciencia*, n.° 47, págs. 58-77; agosto, 1980.

TEORÍA UNIFICADA DE LAS PARTÍCULAS ELEMENTALES Y DE LAS FUERZAS. Howard Georgi en *Investigación y Ciencia*, número 57, págs. 18-36; junio, 1981.

FUENTES TERMALES EN EL FONDO DEL OCEANO

THE MECHANISMS OF HEAT TRANSFER THROUGH THE FLOOR OF THE INDIAN OCEAN. Roger N. Anderson, Marcus G. Langseth y John G. Sclater en *Journal of Geophysical Research*, vol. 82, n.° 23, págs. 3391-3409; 10 de agosto de 1977.

CHEMISTRY OF HOT SPRINGS ON THE EAST PACIFIC RISE AND THEIR EFFLUENT DISPERSAL. J. M. Edmond, K. L. Von Damm, R. E. McDuff y C. I. Measures en *Nature*, vol. 297, n.° 5863, págs. 187-191; 20 de mayo de 1982.

UNA VENTANA ABIERTA AL CEREBRO DORMIDO

SLEEP DISORDERS: INTERSECTIONS OF BASIC AND CLINICAL RESEARCH. Dirigido por M. H. Chase. Spectrum Publications, Inc., 1982.

DIFFERENT BEHAVIORS DURING PARADOXICAL SLEEP WITHOUT ATONIA DEPEND ON PONTINE LESION SITE. J. C. Hendricks, A. R. Morrison y G. L. Mann en *Brain Research*, vol. 239, págs. 81-105; 1982.

PRIMEROS AGRICULTORES DE LA LLANURA DEL NORTE DE EUROPA

EARLY NEOLITHIC SITES AND BRZEŚĆ KUJAWSKI, POLAND: PRELIMINARY REPORT THE 1976-1979 EXCAVATIONS. Ryszard Grygiel y Peter Bogucki en *Journal of Field Archaeology*, vol. 8, n.° 1, págs. 9-27; 1981.

EARLY EUROPEAN AGRICULTURE. M. R. Jarman, G. N. Bailey y H. N. Jarman. Cambridge University Press, 1982.

EARLY NEOLITHIC SUBSISTENCE AND SETTLEMENT IN THE POLISH LOWLANDS. Peter I. Bogucki, British Archaeological Reports, Oxford, 1982.

SEÑALES QUIMICAS DE AGREGACION EN LAS AMEBAS

THE CELLULAR SLIME MOLDS. John Tyler Bonner, Princeton University Press, 1967.

THE DEVELOPMENT OF *DICTYOSTELIUM DISCOIDEUM*. Dirigido por W. F. Loomis. Academic Pres, 1982.

CHEMICAL IDENTITY OF THE ACRASIN OF THE CELLULAR SLIME MOLD, *POLYSPHONDYLIUM VIOLACEUM*. O. Shimomura. H. L. B. Suthers y J. T. Bonner en *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, n.° 23, págs. 7367-7379; diciembre, 1982.

CONTROL GENETICO DE LAS PROTEINAS DEL TRIGO

WHEAT PROTEINS. D. D. Kasarda, J. E. Bernardin y C. C. Nimmo en *Advances in Cereal Science and Technology*, vol. 1, págs. 158-236. Amer. Assoc. Cereal Chem. Inc., St. Paul Minnesota, 1976.

GENETIC CONTROL OF THE CONTENT, AMINO ACID COMPOSITION, AND PROCESSING PROPERTIES OF PROTEINS IN WHEAT. C. F. Konzak en *Advances in Genetics*, vol. 19, págs. 407-582; 1977.

CHROMOSOMAL CONTROL OF WHEAT ENDOSPERM PROTEINS. F. García Olmedo, P. Carbonero, C. Aragoncillo y G. Salcedo en *Seed Protein Improvement by Nuclear Techniques*, IAEA Proc. Series STI/PUB/479, págs. 555-566; 1978.

CHROMOSOMAL LOCATIONS OF GENES THAT CONTROL WHEAT ENDOSPERM PROTEINS. F. García Olmedo, P. Carbonero y B. L. Jones en *Advances in Cereal Sciences and Technology*, volumen 5, cap. I, págs. 1-47, Amer. Assoc. Cereal Chem. Inc., St. Paul, 1982.

TEMAS METAMAGICOS

ANATOMY OF LISP. John Allen. McGraw-Hill Book Company, 1978.

ARTIFICIAL INTELLIGENCE PROGRAMMING. Eugene Charniak, C. K. Riesbeck y Drew V. McDermott. Lawrence Erlbaum Associates, Inc., 1980.

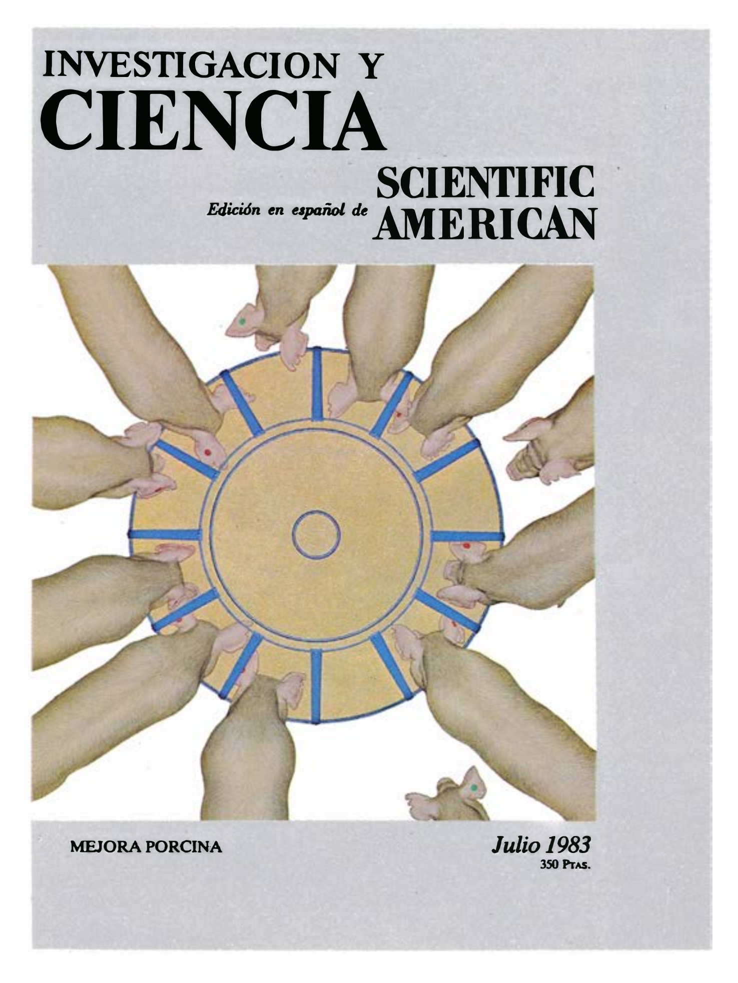
HISTORY OF LISP. John McCarthy en *History of Programming Languages*, dirigido por Richard L. Wexelblat. Academic Press, 1980.

TALLER Y LABORATORIO

OBSERVATIONS OF AN EARLY MORNING CUP OF COFFEE. Vincent J. Schaefer en *American Scientist*, vol. 59, n.° 5; septiembre-octubre, 1971.

GRAIN SHAPE EFFECTS ON SETTLING RATES. P. D. Komar y C. E. Reimers en *Journal of Geology*, vol. 86, n.° 2, págs. 193-209; marzo, 1978.

Seguiremos explorando los campos del conocimiento



ARMAS INTELIGENTES EN LA GUERRA NAVAL, por Paul F. Walker
El impresionante comportamiento de las armas con guía de precisión en la guerra de las Malvinas pone en entredicho la eficacia bélica de los buques de superficie.

VIBRACIONES DEL NUCLEO ATOMICO, por George F. Bertsch
El núcleo puede temblar, resonar e incluso "respirar". El movimiento coordinado de las partículas nucleares nos informa de las fuerzas que actúan entre ellas.

MICROCUERPOS CELULARES, por Christian de Duve
Un grupo de orgánulos subcelulares, unidos por semejanzas estructurales superficiales, incluye varios tipos distintos, cada uno con una dotación enzimática propia.

AVANCES EN LA PRODUCCION PORCINA, por Wilson G. Pond
El cerdo aporta el 25 por ciento de la energía y el nueve por ciento de las proteínas que los humanos obtienen de fuente animal. Factores económicos y biológicos han conducido a un nuevo tipo de cría porcina.

NUEVOS MATERIALES INORGANICOS, por J. D. Birchall y Anthony Kelly
Los materiales orgánicos sintéticos desempeñan un importante papel en la civilización moderna; sin embargo, ciertos productos menos corrientes, fabricados con arena, arcilla y otros minerales aventajan a aquéllos en el menor requerimiento de energía.

MUERTE CARDIACA SUBITA: UN PROBLEMA TOPOLOGICO, por Arthur T. Winfree
Muchos fallecimientos repentinos son consecuencia de la fibrilación, un trastorno de la contracción coordinada de las fibras musculares cardíacas.

METODOS ESTADISTICOS ORDENADOR-INTENSIVOS, por Persi Diaconis y Bradley Efron
A costa de grandes cálculos, es posible prescindir de hipótesis convencionales sobre los datos. Un método en particular ha puesto en cuestión muchas estimaciones de fiabilidad de inferencias científicas.

SIMBIOSIS *RHIZOBIUM*-LEGUMINOSA, por Concepción Azcón González de Aguilar, José Miguel Barea y José Olivares
El incremento en la demanda de alimentos y precio de los fertilizantes nitrogenados ha estimulado las investigaciones sobre esta asociación microbio-planta capaz de integrar el nitrógeno atmosférico en aminoácidos esenciales para la dieta animal.

